Universidade de Aveiro
2016

Departamento de Eletrónica, Telecomunicações e Informática

Luís Pedro Santos de Assunção

Estimação do ritmo cardíaco recorrendo a vídeo no ambito de experiências em psicologia

Heart rate estimation using video in psychology experiments

Universidade de Aveiro
2016

Departamento de Eletrónica, Telecomunicações e Informática

Tese apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia de Computadores e Telemática, realizada sob a orientação científica do Professor Doutor José Maria Fernandes, Professor auxiliar do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro e Professor Doutor Ilídio Castro Oliveira, Professor auxiliar do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro.

o júri

President / president	Augusto Marques Ferreira da Silva Professor auxiliar do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro
vogais / examiners committee	Hélder Filipe Pinto de Oliveira Investigador Senior,INESC TEC, PhD
	José Maria Amaral Fernandes Professor auxiliar do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro

agradecimentos

Quero aproveitar esta oportunidade para agradecer ao grupo de pessoas que sempre estiveram do meu lado, e que graças ao seu apoio, dedicação e confiança me ajudaram a completar este trabalho.

Aos meus orientadores, o professor José Maria Fernandes e professor Ilídio Fernando de Castro Oliveira, pela possibilidade e oportunidade de executar este trabalho, pelos seus concelhos e disponibilidade incansável para discutir, melhorar e ajustar o trabalho, sem os quais não me seria possível levar as tarefas a bom porto.

Quero agradecer a Doutora Susana Brás pela ajuda incansável e paciência, e pelos conhecimentos e indicações na interpretação, organização e apresentação dos resultados obtidos, e a Professora Sandra do departamento de Psicologia pelos conselhos, recursos e experiências anteriores cedidas que permitiram validar este trabalho.

Agradeço também aos meus pais, cuja ajuda, apoio, compreensão e confiança me ajudaram a ultrapassar os obstáculos e me deram força para seguir em frente nos momentos mais difíceis.

Deixo um grande agradecimento e abraço aos meus colegas de laboratório que me proporcionaram um ambiente de trabalho excelente, com todas as condições necessárias, amizade e camaradagem.

Agradeço a doutora Susana, a doutora Raquel, ao professor Ilídio. Ao Pedro Miguel, Pedro Abade e Luis Almeida, por terem aceitado ser cobaias nas minhas experiencias e terem tido paciência e partilharem os seus concelhos e ideias comigo, que tanto me ajudaram a seguir um bom caminho.

Por fim, agradeço aos meus colegas de curso que me acompanharam nesta aventura até ao fim, ao Joel, ao Pedro Abade, Pedro Daniel, Daniel Oliveira, Nicolas, Cláudio, Vítor, Ana Sofia, Pedro Abade, Manuel, Sofia Oliveira, Ana Rebelo. Obrigado pela vossa camaradagem, por me ouvirem e me ajudarem a ultrapassar os momentos mais difíceis, por todas as aventuras e sacrifícios que fizemos juntos, por tudo o que partilhamos, obrigado.

palavras-chave

Eletrocardiograma, ritmo cardíaco, rede neuronal artificial, fotopletismografia, PPG

resumo

Devido aos avanços em visão de computador, vários trabalhos propuseram métodos para estimar o ritmo cardíaco utilizando partes da pele do sujeito (testa, face completa, ...).

Contudo, embora se tenha obtido resultados promissores não há provas conclusivas acerca da precisão e aplicabilidade em condições mais realistas (fora do laboratório), devido aos senários muito controlados ou tempo de amostragem limitado. Nesta dissertação, propusemo-nos avaliar a utilidade da estimação de ritmo cardíaco por vídeo e contruído no estado da arte para atender a cenários mais realistas e exigentes, isto é, cenários menos controlados e avaliar os resultados em sessões de monitorização mais longos (>1 minuto). Nós efetuamos duas experiencias baseadas vídeo de estímulo onde o objetivo era de medir as alterações de ritmo cardíaco produzidas pelo vídeo. Em ambos os senários, ECG foi utilizado para extrair ritmo cardíaco que foi usado como comparação. O primeiro cenário foi adquirido com vídeos cujo estímulo foi “Nojo” (25 minutos), o segundo cenário usou vídeos mais pequenos (<10minutos) usando um estimulo neutro e de “felicidade”. Os resultados mostram que a estimação do ritmo cardíaco é bastante sensível ao ruido e não é clara a relação no estudo completo. Contudo, quando estudamos a relação entre o ritmo cardíaco estimado por vídeo e por ECG tornou-se claro que ambos eram altamente correlacionados em intervalos de tempo limitados, sugerindo que o ritmo cardíaco estimado por vídeo deve ser explorado. Durante o processo desenvolvemos o PsyVidLab que, para alem de incorporar a estimação de ritmo cardíaco por vídeo, permite a aquisição síncrona de vídeo, ECG, e alguns módulos de processamento de vídeo básicos especificamente estimação de emoção de expressões faciais.

keywords

Electrocardiogram, heart rate, artificial neural network, photoplethysmography, PPG

abstract

Heart rate is a relevant physiological marker used in several areas, namely psychology, as a measure of the anxiety and stress among other states. Typically, the heart rate is calculated from ECG that implies using dedicated equipment with electrodes placed on the human subject which can be considered invasive in many situations i.e. not comfortable or humanly suitable.

With the advances in computer vision several works proposed methods to estimate the heart rate from video capturing skin patches of the subjects (e.g. for head, overall face, …). However, although promising results there no conclusive proofs on the accuracy and applicability in more realistic conditions (e.g. outside of the laboratory) namely due to the very controlled scenarios or limited sampling time.

In this dissertation we proposed to evaluate the usefulness of heart rate estimation based on video and built upon the state of the art to address more realistic and challenging conditions i.e. less controlled scenarios and evaluate it under larger monitoring sessions (>1 minute). We performed two experiments based on video stimulus where the objective was to measure the HR changes induced by the video. In both scenarios, ECG was used to extract the HR that was used as ground truth. The first scenario was acquired with videos to elicit disgust (25 minutes), the second using smaller videos (<1 minute) using a neutral and "happiness" inducing videos.

Our results show that the heart rate estimation is very sensitive to noise and not clear relation on the complete studies was observed in any of the scenarios. However, when studying the relation between the HR estimated from video and from ECG it was clear that both were highly correlated in limited time intervals suggesting that video estimated HR may be worthy to explore.

In the process we developed PsyVidLab that besides incorporating the video estimated HR allows synchronous acquisition of video, ECG and some basic image processing modules namely emotion estimation from facial expression.

I. Contents

II. List of figures

III. List of tables

1 Introduction

Psychology is a field of study that requires a lot of interaction with people, be it in research or in therapy. Video resources plays an important role in several psychological studies, including in research [1].

In Psychology, most diagnostics and interpretations rely on visual observation [6] [7], however it was found that only one-third to one-half of the results in psychological studies results were reproducible [8], thus strengthening the need to have means of acquiring data that does not depend on the subjectivity of the observer. Recorded videos may offer an option as they allow the review of the experiments to evaluate behaviours and context alterations that might prove relevant [2], and can be used in computer-based analysis methods.

In the past few years, several advancements in computer vision offered a new opportunity especially when supported in computational methods that provide reproducibility and more precise solutions. Computer vision has been used to perform measurements of biometric data based on image, e.g., facial recognition [3], and heart rate estimation from video [4]. It has already been used to measure physiological variables, namely heart rate –a commonly used measure to assess human physiological status. Several studies show that we can effectively estimate heart rate and heart rate variability from video [4] [6], and that the potential to use video as a quantified and complementary source of information in characterising the behavioural and biometric components of subjects is definitely present.

Using computer vision techniques allows more flexible solutions such as telephone cameras or action cameras that can be used and easily deployed in more challenging scenarios. Video has also the advantage of being a non-invasive option (i.e. no need of contact of electrodes or probes).

However, when using video there are two main aspects to consider: 1. Often times the setup is installed on a laboratory restricting their use, and 2. Rarely does the information contained in the video beyond that of control is explored, that is, to allow reviewing the experiment as an external observer.

1.1 Motivation

Currently most studies performed in psychology rely heavily on qualitative observations that are naturally subjective, conditioned by the way the observer perceives the environment, namely by their motivational states, wishes and preferences [5].

Psychologists sometimes struggle to find the required tools to assist them the way they want on their experiments or consultations. Some solutions exist such as Biopac [24], but are very expensive, and limited to the laboratory they are installed on. In addition, they require probes and electrodes that may interfere with the experiment, and may even upset or distract some patients (such as those suffering from Autism). They also may have video sources from past experiments that they might want to extract relevant information from, such as heart rate.

This work's motivation was to explore the use of video as a non-invasive option to measure physiological variables within an experimental setup deployed outside the laboratory, by developing a proof of concept or technological preview of a system capable of such features. Avoiding the use of electrodes or probes, helps the observed subject to be less aware of the observation process and react in more natural and undisturbed way.

Even though we mention psychology as a use case, such system is not limited to a particular need, and can be used on other circumstances such as a hospital room. We do not intend to develop a system that will compete with or replace current solutions, instead we will evaluate the relevance and potential of these methods for the future and prove that they can in fact be used after further development and honing.

This dissertation work was developed in collaboration with the Education and Psychology Department in University of Aveiro, with Professor Sandra Soares, to lay down the foundations that could stimulate development of technologies capable of supporting experiments inside and outside a laboratory setting and to re-evaluate past experiments that were recorded in video. Using computer vision techniques could help extract quantified information from those sessions recordings. This served as our problem context, and it is why we mention usage in psychology several times during our work.

The opportunity to present other sources of biometric data was also identified, given the relevance of the heart rate in the physiological assessment of the subject, our focus was on evaluating the role of video as a heart rate estimation tool – both as online monitoring tools and as offline tools over past trials. If video proves to offer a valid option, in some settings, it can provide a good trade off solution for estimating HR in scenarios where invasive solutions are not applicable or may induce unwanted response from the subjects.

By using video to estimate heart rate, the potential users of this technology would have freedom to leave the rigid environment of the laboratory and setup their experiments where they want, without troubling or distracting the subjects with preparation and installation of electrodes.

Furthermore, certain types of experiments or subjects require that no electrodes or probes are used so that the outcome is not affected and the experiment is not disturbed, for example in the case of autists which are oversensitive.

1.2 Objectives

With this dissertation we aim at developing a system to support psychophysiological data acquisition and analysis capable of:

- Extracting relevant biometric information (in the form of heart rate) from past recorded experiments
- Acquire and record relevant biometric data from live experiments in a contactless un-intrusive way (without electrodes, in the form of heart rate)
- Support traditional biometric acquisition methods, which in this case would be ECG (this would allow us to test our estimated heart rate against proper readings)
- Recording the experiment (the subject and its reactions) in video, and the possibility of having multiple ways of doing this concurrently (for example having two cameras)
- Present stimulus in the form of video
- Previewing the various sensors, and the ability of viewing sensor data in real-time
- Supporting an event system in order to tag relevant observations during the experiment

- Storing the experiment results in terms of the data it produced

The system is named PsyVidLab to reflect its ability to record multimodal acquisitions, featuring heart rate estimation and video recording. It does not take in to account usability, and merely serves as a proof of concept of the system, a technology demo aiming at verifying the usefulness of the techniques used. The results of the methods used will dictate the interest in developing a fully fleshed out program.

Besides the system features, another goal is related to quality assurance and system validation, including:

- Evaluate if the heartrate estimation has a good quality (have a difference less than 10 beats per minute with respect to the reference measurement)
- Validate if the application acquires data correctly (specifically ECG), and that the various signals are synchronized
- Verify if the application is capable of running in real-time

It is important to have the means of visualizing the data that our system produced. This review application would feature the ability to perform heart rate variability processing on ECG signal, and display the various signals in the same common time (the experiment) synchronized with the recorded videos and events.

2 Background Concepts and State of the Art

In this section, initially we will address the use of video in estimating heart rate and emotion. Both solutions can provide valuable solutions to be used in the psychology assessment and studies.

2.1 Estimating Heart Rate from Video

Heart rate is an important variable used in psychology, as heart rate changes when a human subject receives external stimuli [4]. Although heart rate can be measured using electrodes, this approach is often too invasive and may influence the heart rate measurements. So using non-invasive solutions (i.e. without probes or electrodes), in case they exist, is a welcomed option. One of these non-invasive options is based on photoplethysmography, also abbreviated PPG. PPG consists in measuring the light reflected from the subject’s body that carry information about subtle colour variations [9]. Colour changes due to blood flow are generally very small and only though video processing techniques can we amplify those variations. The most notable example is Eulerian Video Magnification (EVM) [34].

The usefulness of these measuring techniques opens the possibility of using video beyond a review perspective and use it as another source of information, and increase the flexibility to use outside the laboratory. Some work has already been done in this area, however it was never used in a psychology setting for extended periods of time, to our best knowledge.

The several available PPG methods are based on a generic computational image processing sequence (as depicted in Figure 1). A frame is acquired using video/camera. The frame is then segmented to extract the region of interest, like a forehead or a wrist. Then the region is processed and analysed from colour changes in specific spectrums, and a heart rate is estimated based of the frequency at which the colour changes or oscillates.

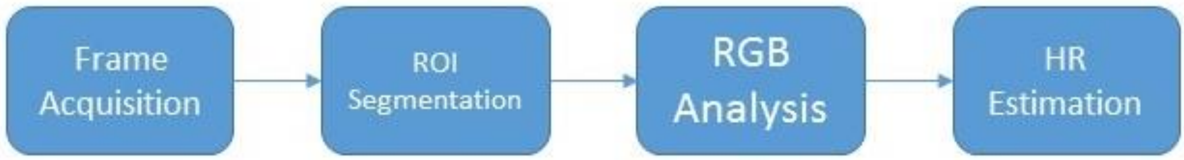

Figure 1 – Generic method for photoplethysmography

The different PPG methods typically have three main concerns on handling:

- Subject Motions,
- Ambient Illumination
- Sensor Spectrum Sensitivity.

Table 1 summarizes contributions in photoplethysmography regarding motion tolerance:

Table 1 - Photoplethysmography contributions by author on Motion Tolerance

Work	Main topic addressed
Wieringa et al. [10] (2005)	Focused on rigid, stationary Region of Interest, used monochrome CMOS cameras with apochromatic lenses and LEDs to measure reflected light from the wrist at three specific wavelengths (660nm, 810nm, 940nm)
Poh et al. [12] (2010)	Images the subject’s face, and allows limited and natural motion, like slight oscillation of the head and occasional changes in posture
Cennini et al. [13] (2010)	Estimates the contribution of motion artifacts and corrects PPG signal using adaptive filtering
Sun et al. [14] (2010)	Focused on the user’s face, but with strict limitations in motion
Estepp et al. [15] (2014)	Accounts for head rotation as opposed to translation, by using a multi-imager array with blind source separation to reduce the effects of rigid head motion artifact in the measurement of pulse rate using PPG
Haan et al. [16] (2014)	Proposes a new approach to PPG based on linear decomposition of color channel information dictated by optical properties of the imager setup and taking in account the illumination source spectrum
Chung et al. [19] (2015)	PPG method that allows for motion tolerance by using motion information from location tracking.

In terms of illumination tolerance, the following authors have produced interesting results (Table 2):

Table 2 - Photoplethysmography contributions by author on Illumination Tolerance

Work	Main topic addressed
Amelard et al. [20] (2015)	Supports dynamic ambient lighting conditions by making use of a temporally-coded light source and synchronized camera to compensate for light changes

It is assumed that in most use cases (laboratory, indoors) the ambient light is relatively constant, however when outdoors, viewing a video or in virtual reality, the ambient light might variate. It has been shown that changes in illumination affect absolute, and not relative, magnitude of the PPG waveform [14], however its effects on current PPG methods are unknown. Furthermore, the qualities and properties that constant lightning needs to have to produce quality PPG waveforms are undetermined [27], and we believe that further experiment and exploration in this area could be of interest.

Table 3 lists selected contributions with respect to sensor spectrum sensitivity:

Table 3 - Photoplethysmography contributions by author on Spectrum Sensitivity

Work	Main topic addressed
W. G. Zijlstra et al. [35] (1991)	The green and orange visible spectrum carry the most information for oxygenated and non-oxygenated blood
D. McDuff et al. [36] (2014)	Five-band visible camera with cyan, orange, red, green and blue yielded good results when combining the orange, green, and cyan channels
N. Garbey et al. [37] (2007)	Infrared cameras have been tested and proven capable to provide results when used to measure temperature fluctuation on the carotid region to calculate the pulse rate
Verkruysse et al. [11] (2008)	Used a regular RGB camera (like webcam) for acquisition.

Contributions to PPG previously shown make use of the RGB spectrum, however, exploring different spectrums and combinations can yield better results as shown by D. McDuff et al. [36].

Our focus was on the usage and integration of a method available. Therefore, we chose to use the contribution of Thearn [25] which presented a method of real-time heart rate estimation, which we altered to meet our specifications. His contribution had similarities to Wieringa et al. [10] since it used a fixed region of interest, which we altered to a moving region of interest by continuously detecting the face of the subject; and similar to Verkruysse et al. [11] in that it used a regular webcam (RGB spectrum) for acquisition, focusing on the green channel. We investigated the motion tolerance method described on table 1, however we chose to keep the method simple and verify the results it would provide. If the results are good enough, then the basis for further improvement are set. In terms of spectrum sensitivity, from our readings in N. Garbey et al. [37] the infrared spectrum showed great promise, and we would have liked to explore it, however we didn’t had access to the right hardware, and we considered the RGB band far more easy to handle and acquire in the market (true infrared cameras are very expensive).

The contributions summarized in Table 1, Table 2 and Table 3 tested their methods with small experiment durations, around 30 seconds. With these small timeframes, their results were good and provided a reliable estimation of the heart rate. While some solutions are hard to deploy in real world scenarios, such as N. Garbey et al. [37] and Amelard et al. [20], the other contributions suggested solutions that are easy and capable of being deployed and used in real world usage, which is what we desire.

2.2 Emotion Recognition from Video

Regarding emotion recognition, which is a useful metric in psychology, for example used in the detection of some illnesses such as schizophrenia [38], several methods exist that make use of video processing techniques.

The following table summarises the contributions we considered the most relevant:

Table 4 - Emotion Recognition from Image contributions by Author

Work	Main topic addressed
L.C. de Silva et al. [21] (1997)	Multi modal approach by making use of both visual and auditory data to recognize the subject's emotion
N. Sebe et al. [22] (2002)	Makes use of Cauchy Naïve Bayes classifier to recognize emotions in video sequences
Spiros V. Ioannou et al. [23] (2005)	Uses fuzzy neural networks and facial animation parameters present in MPEG-4 to create a robust facial analysis system. Does not require training, but requires a set of linguistic rules
Singla, Rajneesh [39] (2011)	Uses PCA (Principal Component Analysis) and Fisher face algorithm to perform the recognition
Thuseethan, S. et al. [40] (2016)	Makes use of PCA (Principal Component Analysis) to extract the eigenface images. Euclidean distance between the extracted eigenface images and the dataset is used to detect the emotion
Van Gent. P. [41] (2016)	Detects facial landmarks, and makes use of a linear Support Vector Machine (SVM) previously trained with a dataset to predict emotions

When searching for the state of the art on emotion recognition, we wanted current technologies that were easy to understand, implement and adapt to our specific needs, since all we required was a method that could solve our need for a real-time emotion recognition.

Commercial tools include Emotient [46], Affectiva[47] and EmoVu [48] which provide API to embed their methods and algorithms out-of-the-box.

3 Use Cases in Psychology

From the inception of this work the main question was: How to perform heart rate readings in the least intrusive way? This question arose from our collaboration with the Education and Psychology Department in University of Aveiro, where such video base solution, if possible, would be a valid contribution to their experimental setup in this context. One of our initial concerns was to understand where and how the video could be of help in their experimental protocols.

The domain requirements pointed to an application to support the experimental setup, including:

- Configuration of the experiments: with sensors, stimulus to be used
- Preview sensors data
- Management of the experiment: start / stop multimodal recording
- Tagging relevant time events to assist in experiment review
- Present visual stimuli at specific moments

The previous points refer to actions that would be desirable to take on our application, and as such can be seen as a use case each.

Another conclusion is that the application should be experiment agnostic i.e. try to avoid requirements specific to particular experimental protocols. This could help ensure that the application could be applied in different experiments, with different sensors and with different objectives such as emotion recognition, attention (e.g. gaze tracking) or more generic behaviour analysis (e.g. recording subject’s responses).

Ultimately, an application that would include these features was envisioned, that would feature the capacity to change which sensors would be used, the ability to perform the readings in real-time and show the state of said sensors as well to the psychologist during the experiment. An event system that would create temporal marks on the signals produced would have to be implemented as well, since if the psychologists noticed something useful, instead of taking note, they would simply mark it directly on the physiological signal facilitating posterior review. The ability to include the means of displaying the stimulus from the application itself was required also, as keeping all the features in one place would make the live of the psychologists much easier. Figure 2 refers to the uses cases that our application would support.

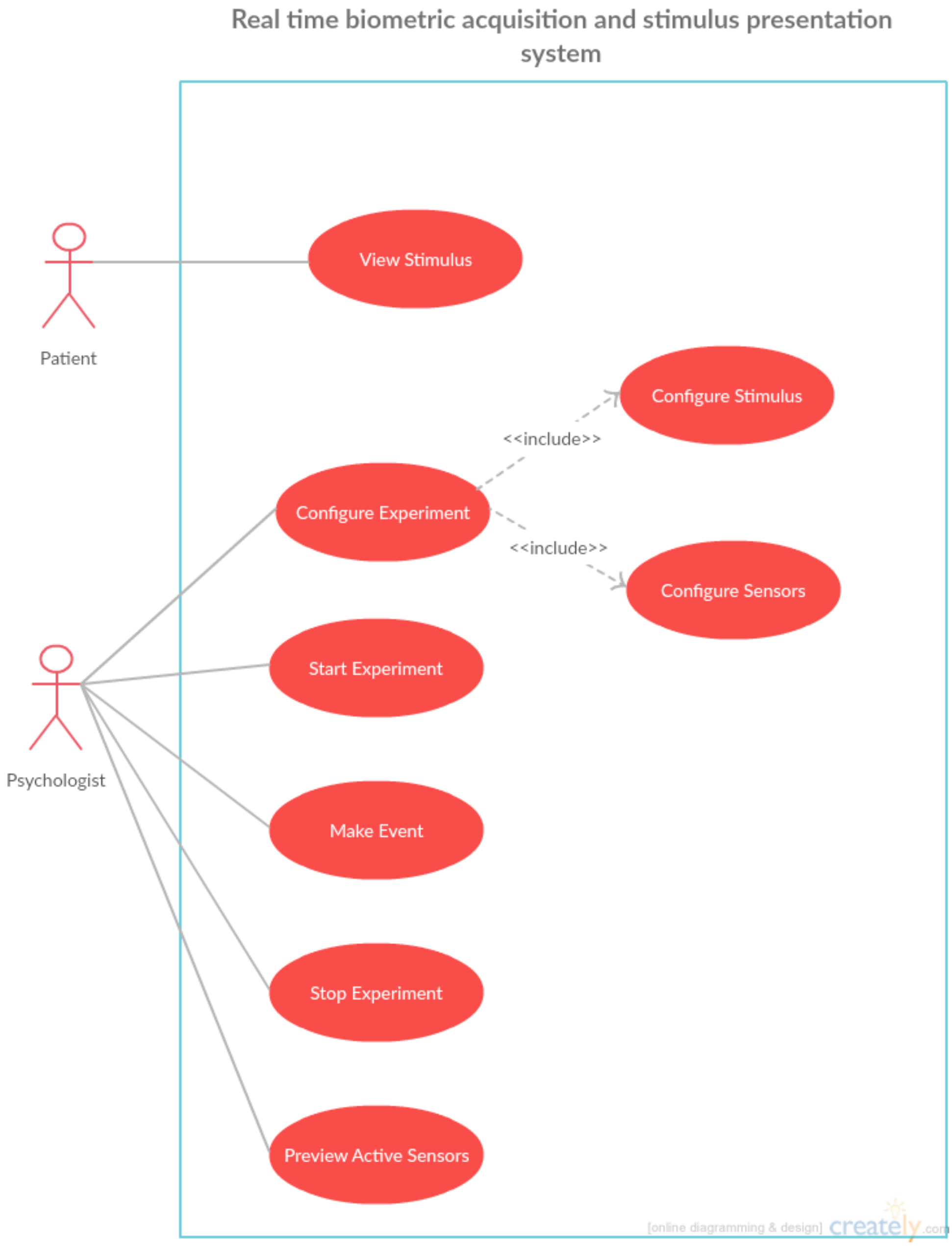

Figure 2 – Use case diagram of the biometric and stimulus presentation system

Although the use case has the psychologists in mind, the system can be used by different entities as long as and the goal is the same: Record physiologic responses without the need of using electrodes in real time, in response to a stimulus.

3.1 Extracting heart rate from video scenarios

Given our focus on video based solutions several scenarios were envisioned and considered when planning the features, flows and development of the system. Two base scenarios were immediately identified, which are: 1). Extract heart rate from recorded experiments (offline), and 2). Perform heart rate readings from on-going video experiments (online).

Regarding the first scenario, extract information (heart rate) from recorded experiments, the rationale involved was: Given that we have a video to analyse, it becomes important that we develop an application capable of reading the video files, pass them to the heart rate estimation module, and store the results as a signal. Therefore, only the heart rate from video module is required, and the subject in the recorded video needs to be frontal facing to facilitate detection of the face.

Scenario 2, perform heart rate readings from video in experiments, is more complex than the first one in terms of requirements. The subject must be placed in front of the camera used to acquire the video in order to facilitate detection of the face. The subject should have its own screen to view the stimulus, and the psychologist his own monitor where he can observe the biometric signals, in this case heart rate, with the added ability of being able to mark the signal, with a brief description, to facilitate posterior review. The stimulus can be displayed from within our application, or rendered externally.

Other scenarios should be supported, using the scenario 2 as its foundation. Instead of using only the heart rate from video module, we could use the ECG module to record ECG data, the gaze tracker to record heat maps over time, and the emotion recognition module to detect emotions as the experiment progresses. These modules, or sensors, can be used at the same time, or choose individually which one to activate. This allows the possibility to adapt the experiment setup to the user's requirements, with multiple combinations. By using multiple signals, for example heart rate from video, ECG and emotion recognition, the user will be able to correlate these multiple signals with each other and specific time instances on the stimulus video, and the recorded video of the subject, in order to find hidden relations that otherwise would not be as easily or clearly visible.

4 PsyVidLab Solution

PsyVidLab is the main contribution of this work. The name consists on the portmanteau of the name psychology, video (since it focuses on video processing technologies), and laboratory (because our system comprises a collection several methods). Again the term psychology is used due to our collaboration with the psychology department, however it is not limited to psychology usage.

4.1 Rationale

One of the initial objectives was to provide a tool that could prove the potential of video processing techniques to help psychologists performing their experiments, ideally without the need of dedicated hardware, and that could be used in and outside the laboratory. Key options were already established related to exploring the potential of computer vision techniques:

- Use video processing technologies for heart rate estimation
- Recognize emotions from videos for automated video segmentation

Specifically, we idealized our system dedicated to assisting and acquiring biometric data during experiments, to contain the following technical features:

- Include the means to estimate the heart rate in real time using video feed from a webcam.
- Recognize the emotion in real time using the video feed of a webcam.
- Previewing the output of the sensors to help detect and solve problems related to detection and performance.
- Must have the option of recording electrocardiogram by using an inexpensive board that makes use of electrodes, for example, Bitalino [29].
- Support the ability to record the experiment using a webcam.
- Communicate wirelessly with a portable action camera to perform recordings from a different angle in necessary. We chose the GoPro due to the out-of-the-box readiness and the very high quality of the images in terms of resolution and temporal frequency.
- All sensors must be modular, and there for only the ones we need should be activated.

The real-time capability was deemed necessary as it allows the user to check the state of the sensors as the experiment unfolds, meaning that in case there is a problem with the readings, it can be detected in time and solved. A webcam is a generic device, as the drivers used by these devices are shared among multiple devices. We chose this approach in order to increase the compatibility with a wide range of devices.

We chose Bitalino as the means of acquiring ECG through dedicated hardware because it is relatively inexpensive, and easy to use, while supporting a good sampling rate (1000Hz). GoPro was the natural choice for an alternative way of recording, given it supports wireless control, good battery time, and good image quality. Even though it’s a relatively expensive equipment, it’s widely available in the consumer market.

The following components are expected to be a part of PsyVidLab:

- Video acquisition module: This module is able to receive, store and provide to other modules video information acquired during trials. The video sources could include from standard cameras, web cams to action cameras like GoPro.
- PPG module: The PPG module is in fact a video processing module that is responsible for processing the video and extract PPG related values namely heart rate estimation from video.
- Physiology monitoring: Although our focus is on non-invasive techniques, our architecture must integrate other sources of physiological monitoring
- Data management: Responsible for storing and retrieving the experiment information
- Synchronization module: This module is responsible for providing a common time reference to all incoming data and events to ensure that later on is possible to build a time coherent overall scenario of the trials.
- Event module: Handles incoming events – trial related (start / stop) or experiment related (custom event)
- Stimuli presentation module: This module provides the resources needed to present stimulus in the form of video, and choosing the appropriate way of doing that
- Configuration module: Provides interfaces for adjusting the settings of the different modules namely experiment related (e.g. stimuli configuration) to more device specific (e.g. video sampling rate, which sensor will be active)

4.2 System architecture of PsyVidLab

Videos are widely used to perform experiments and stimulate reactions on patients. Our program (PsyVidLab) will support the ability to display the stimulus video in full screen while the data is recorded on the background in real time. If two monitors are available, then one of the monitors could display the video and the other would display a dashboard showcasing in real time the various biometric.

We consider the method of estimating heart rate from video a sensor hat would support the input either from a live webcam feed, or from a video, without distinguishing the type of input.

During experiments, we often felt the need to mark, or timestamp, certain moments in the experiments that we thought might be relevant during analysis. These timestamps would also facilitate synchronization between the various signals also and were deemed as necessary as they would benefit the examiner when reviewing the data.

The general architecture of the system is presented in Figure 3, further discussed in the next sections.

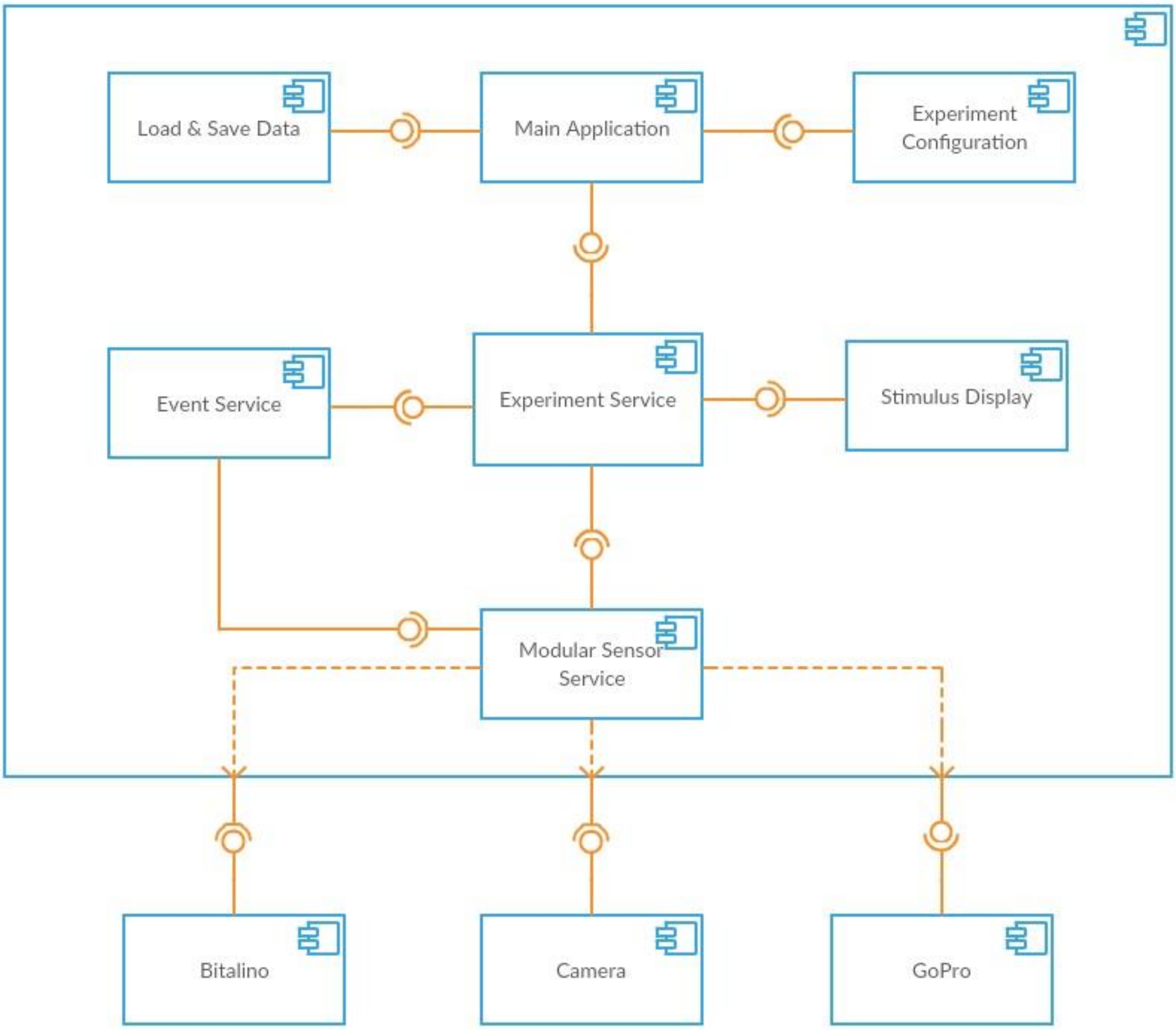

Figure 3 - PsyVidLab architecture

4.2.1 Modular Sensor

This module interfaces with the hardware necessary to acquire data, such as Bitalino, camera and GoPro. Due to the modular nature of the system, A subset if the sensors can be selected and activated. Depending on the configurations the user selected, this component will instantiate the required sensors and their logic, and maintain the data structures associated with the input and output of the sensor modules. All sensor modules are capable of emitting and receiving events, and in conjunction with the Event Service Component, will wait for each other to be ready to start operating simultaneously to guarantee a synchronized start. This module is also responsible to use the computer vision techniques for heart rate and emotion recognition when necessary. The components we described earlier, which are video acquisition module and PPG module are part of the modular sensor service.

4.2.2 Event Service

The Event Service functions in conjunction with the Modular Sensor Service and the Experiment Service collecting and processing their events. It stores relevant events and when they occurred on the biometric signals produced by the Modular Sensor Service and the Experiment Service, with proper descriptions and codes to allows us to differentiate between different occurrences. Events can be automatic such as those emitted by the Modular Sensor Service when starting an experiment, when all sensors are ready, or when a certain threshold on the readings has been reached; and they can be produced manually by the Experiment Service with the user, to mark relevant occurrences during the experiment.

4.2.3 Experiment Service

During the experiment session (and acquisition session), the user has the ability to overview in real time the sensor states from the activated sensors. This module supports that overview, start and stop of the experiment, and the emission of events manually generated by the user to mark relevant occurrences on the biometric signals.

4.3 Heart Rate from Video Estimation module

The heart-rate-from-video module operates on video frames, that means that we can estimate the heart rate in real time from a webcam, or from a recorded video. Psychologists that don't have the required hardware at the time they are conducting experiments, may record the experiment in video to help them on posterior review. Our module can take those videos and estimate the heart rate thus solving the possible lack of hardware, and providing an additional source of information.

The heart rate from video algorithm implementation can be described by Figure 4:

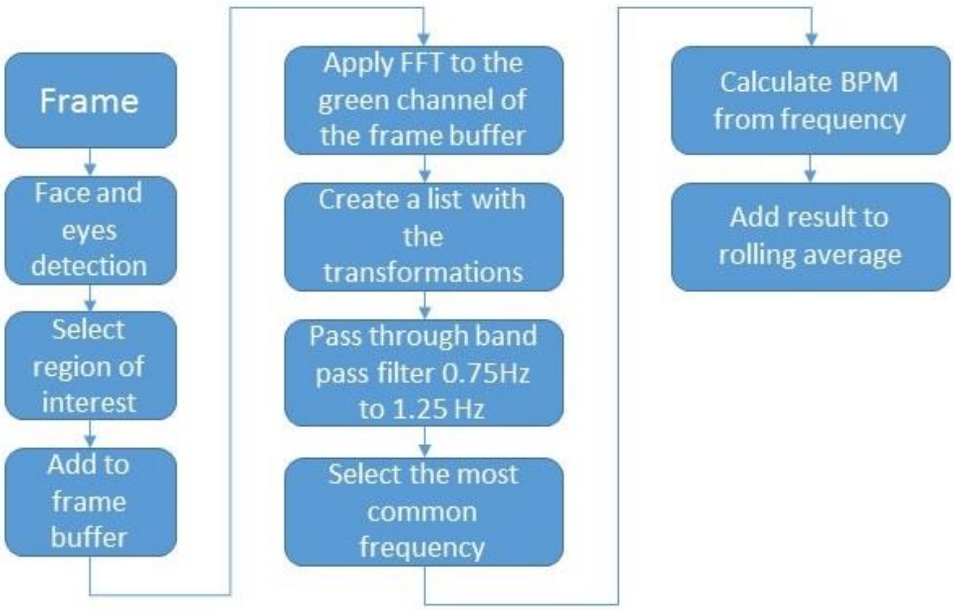

Figure 4 - Operation cycle describing the method used to estimate heart rate from video

This module can be used under extended periods of time. It requires an initial time to stabilise the results since it uses a buffer. Ideally it should be better if the subject remains still or at least with little movement, since too much movement can introduce noise to the signal. Ambient luminosity can also make a significant difference, it's advisable to keep luminosity stable during acquisition.

We modified the solution proposed by Thearn [25] to fit our requirements, specifically during capture our version tracks the user at all times, while the original solution forced the user to stay fixed in one place, it would not track and instead perform the measurements on a specific fixed location of the frame. We also filter occasional spikes by using a rolling mean on the outputs while the original method simply outputted the values without smoothing. In order to fit our programing paradigms, portions of the original code were rewritten to be object oriented.

4.4 Emotion Recognition Method

The emotion recognition module based on an existing method [28] has not been included in the implemented system, however it functions as a standalone application. The way it operates can be described by the following figure:

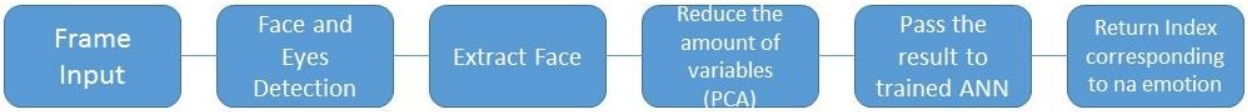

Figure 5 – Operation cycle showing the emotion recognition algorithm we used to develop the method

In order for the module to work properly, the Artificial Neural Network needs to be trained with a dataset. The more types of faces and expressions the dataset has, the most like the module is to detect correctly the emotional expression. Training must be done before the module starts functioning, if enough faces and expressions are added to the dataset, then the artificial neural network only needs to be trained once and keep functioning with the same dataset.

4.5 System Architecture of Review Application

All this data needs to be presented in a way that could be easily interpreted by the psychologists. Another program as idealized that would take all the experiment data and present it in an interactive way, thus simplifying the professionals work by arranging all data in one place. Each module helps solve a particular problem or requirement; the heart rate from video solves the need to have heart rate estimated measurements when there is no dedicated hardware to that end available, when a non-intrusive approach is desired, or when probes or electrodes can´t be used; ECG from Bitalino solves the need to have precise electrocardiogram data by making use of electrodes, but at a very cheap price when compared with Biopac for example; the emotion recognition module provides a way to detect and assist professionals in determining what emotion the subject was experiencing; both recording from webcam and GoPro help solve the need to have a recording of the experiment or session; and the event system helps tag signals and data with timestamp that help the professionals remember and pinpoint the location where something interesting happened during the experiment, as well as helping synchronizing the various signals captured; and finally the stimulus presentation module helps integrate the experiment itself with the measuring tools.

Regarding the second application, in order to have an overview of the experiment that would allow easy interpretation, we chose to plot the various biometric signals together on a timescale, where the beginning of the signals would coincide with the beginning of the experiment, and the end of the signals would coincide with the end of the experiment. The events that the examiner would launch would appear on the correct time marked on the plotted signals. The videos recorded during the experiment would be able to be played alongside the plotted signals and as the video progresses, a mark on the plots would indicate where that moment in the video would relate to the signals.

In summary, the second program, the reviewing application, would have to:

- Be able to plot the biometric signals in regards to the duration of the experiment.
- Signals would have to appear synchronized and tagged with the events timestamps.
- Videos recorded should be playable and have a direct relation between its time and signal instances

The reviewing application would also allow the analysis and viewing of electrocardiogram signal if available, such as heart rate variability measures. The process involved on these operations are heavily based on the work done by Paul Van Gent [30] [31] [32].

Figure 6 illustrates the architecture of the reviewing application.

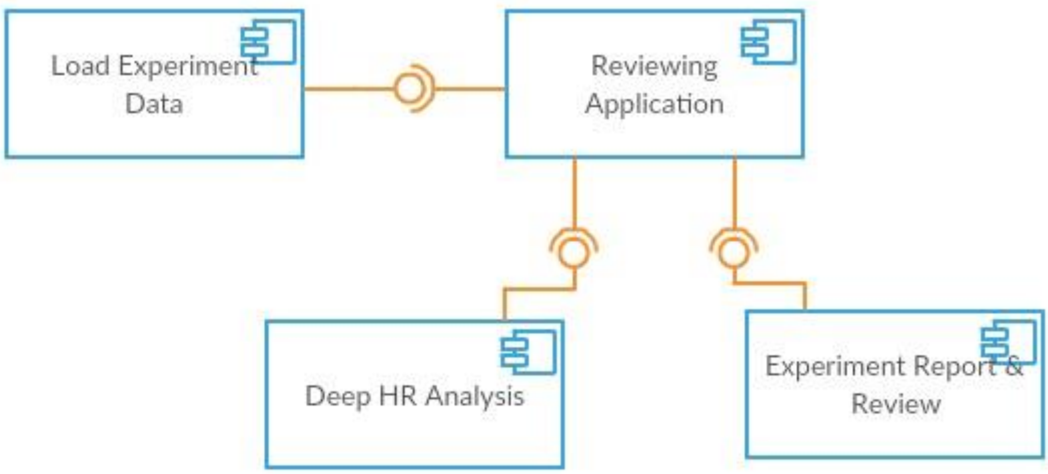

Figure 6 - Experiment overview application architecture

4.5.1 Deep HR Analysis

This module allows in depth analysis of ECG when available in the experiment. The method features heart rate variability measurement, such as SDNN, SDSD, RMSSD, NN20, NN50, pNN20 and pNN50 besides the actual heart rate. These metrics have proven to be useful in psychophysiology, for example allowing detection of anxiety [42]. Furthermore, this module will draw a graph with ECG and beat detection to facilitate further analysis of the signal by psychologists.

4.5.2 Experiment Report & Review

This module will take all the signals acquired during the experiment, and draw them overlapped while keeping scale correct. Since the signals are synchronized, they start and end at the same time, and have the same time axis. Recorded events will be drawn with the signals specifying the moment where they occurred and their description. Video sources from GoPro or webcam are also synchronized, as well as the stimulus video. Playing these video will create a mark on the graph that tells us where the moment on the video correspond on the signal.

5 PsyVidLab implementation

In this section, we will address the implementation of our system, showcasing the UI of our proof of concept psychology workflows we took under consideration when developing the system, validation of key strategic components, technical difficulties and their implemented solutions.

Implementation overview of our current implementation contains the following modules:

- Two video acquisition modules that support both webcam and GoPro
- One bio signal acquisition module using Bitalino to perform ECG capture
- One PPG (heart rate from video) module making use of the algorithms proposed by Thearn[25]
- A stimulus presentation module capable of displaying video
- An event module responsible for tagging timestamps though the experiment

Our solution is a modular system with a frontend application, where the user can activate or deactivate the modules to conduct the experiment in the way he requires.

The following figure shows both the main page of the application and its general configuration page.

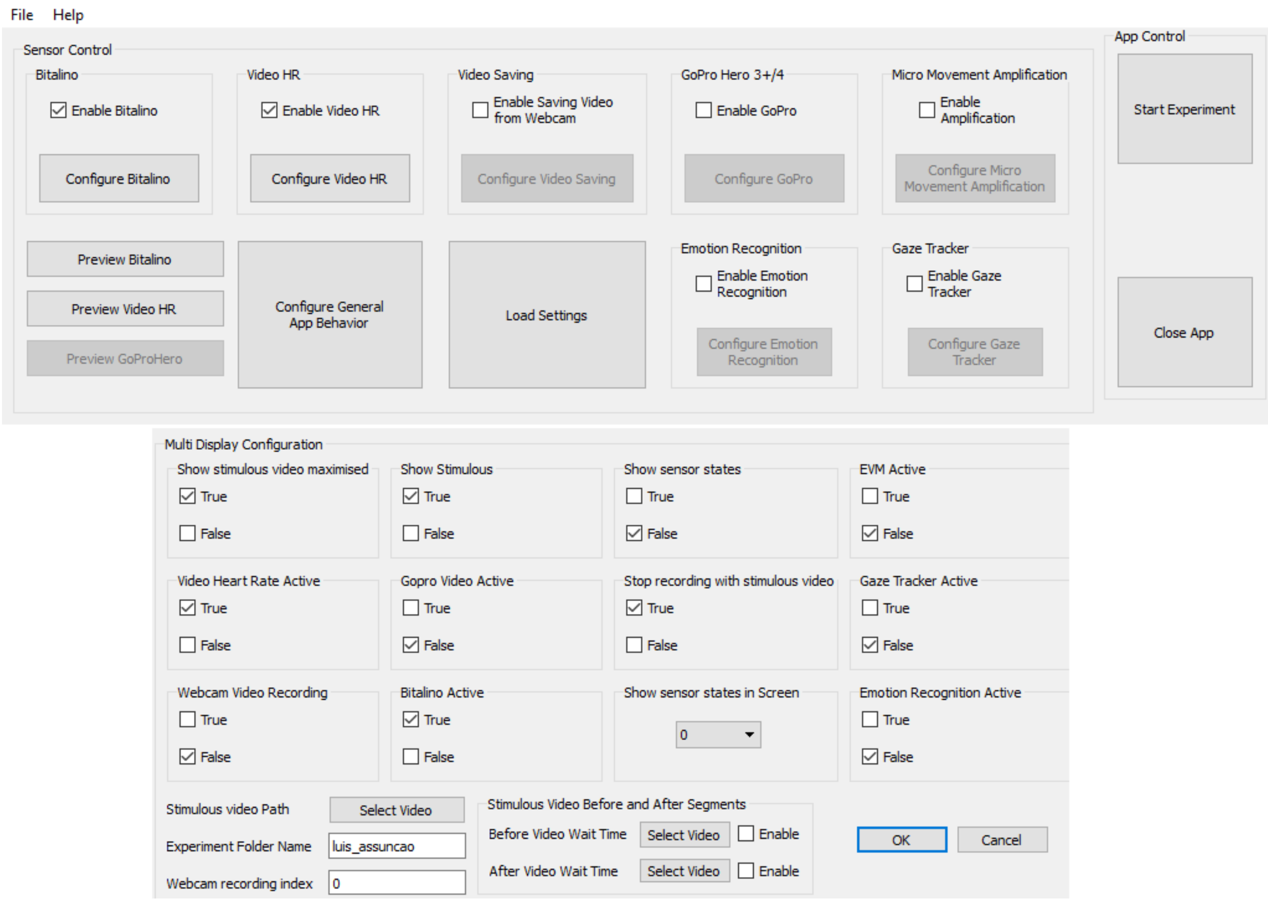

Figure 7 – PsyVidLab configuration view

Figure 7 presents the main interface that allows selection sensors to be used / available by using a checkbox to be activate them. Bitalino and Video HR module (estimate heart rate from video) are considered as individual sensors although the first provides several readings and the second, in fact, is a processing module that provides an HR estimation based on video. Several modules were implemented but not integrated in the current version as Micro Movement Amplification, Gaze tracker and Emotion Recognition. The Configuration page allows setting some common protocol related customization namely configuring the video stimulus (if any) or establishing the baseline duration in the experimental protocol.

Once the experiment is running, the application shows the dashboard in Figure 8.

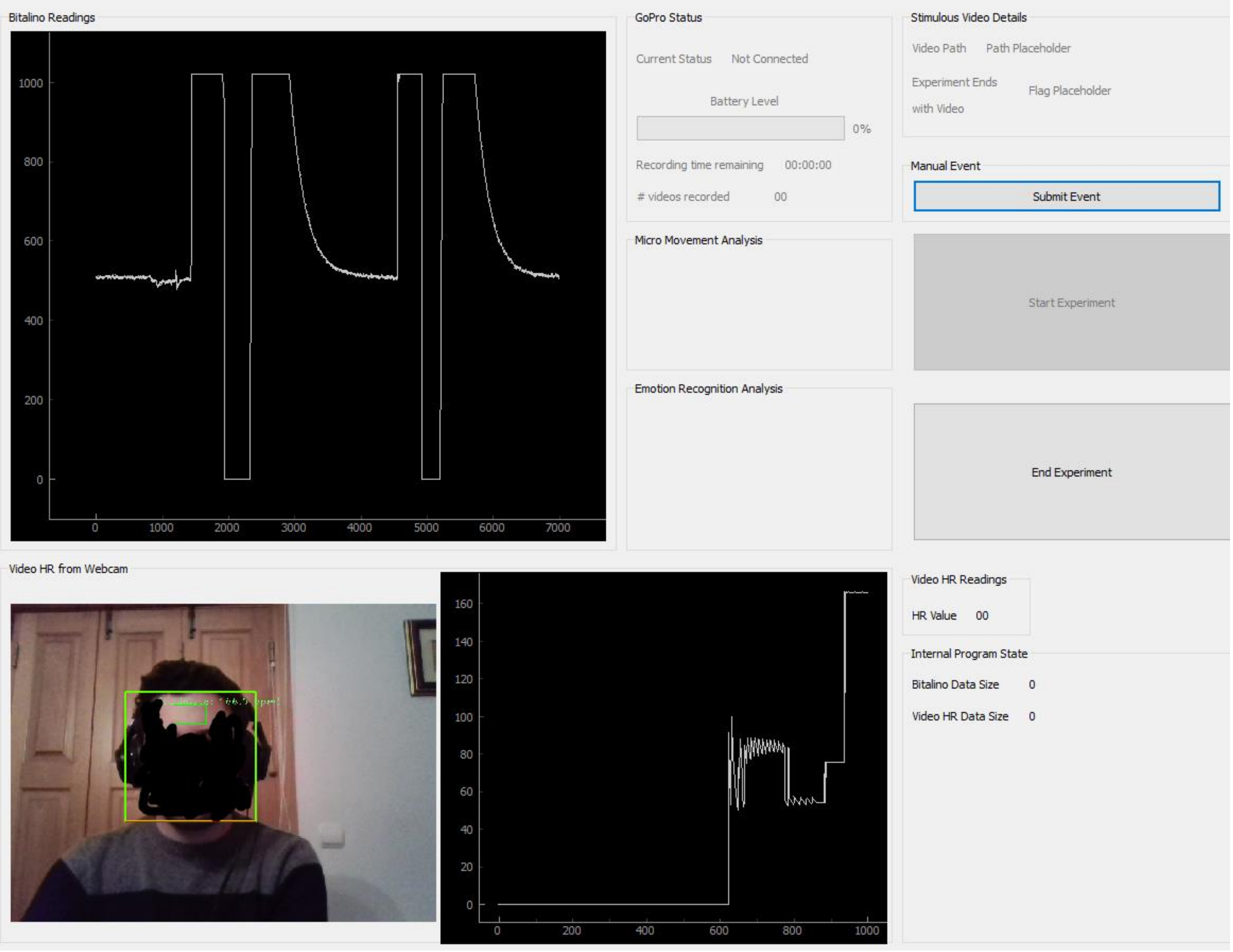

Figure 8 - PsyVidLab dashboard during an experiment

The dashboard features two real-time graphics showing the ECG from Bitalino and the estimated heart rate from video, along with the video feed from the webcam. If the heart rate from video module is activated, this video feed displays the detection of the face and the region of interest used in the estimation, otherwise it shows an unaltered video feed. A Submit Event is present which allows us to tag signals with a description in time and previously explained, and certain controls were placed in anticipation for future use and development, such as Micro Movement Analysis and Emotion Recognition Analysis.

5.1 System workflow

We will describe the workflows of our system that better explain the sequence of actions needed to perform certain actions such as start experiment and preview sensors. The following diagrams illustrate the typical usage scenarios, and how they can be accomplished to perform certain tasks.

5.1.1 New Experiment Setup

In order to start a new experiment, first we need to configure it to our needs. That involves choosing the data we want to acquire, for example heart rate from video, ECG, webcam recording, or GoPro recording.

We can use our application to show the stimulus itself, or use an external source for stimulus. After choosing what our experiment will feature, the time to try out the configurations before the experiment comes to see if all configurations are correct and according to our needs, if any sensor or peripheral is malfunctioning, or if the operational circumstances of the experiment are correct, that is, if the user is centered on the video, if the electrodes are well positions, if the illumination is adequate, among others.

To assist the systems user in that task, we provide real time visualization of each activated sensor, to help detect and solve the operational mishaps that may occur, and if remained unchecked, could slip thought the experiment and jeopardize the quality of the study when analyzing the results posteriorly.

After these pre-experiment checkups, it’s time to start the experiment itself if the user so desires. Once the experiment starts, the user still has an overview of the states of all activated sensors so that if there is any problem, it can be detected and resolved, or at least taken note of. The event system can help mark interesting happenings during the experiment, and can be also used to mark in time on the signals where a problem occurred.

After the experiment is finished, or when the user so desired, the experiment ends, and acquired data and events are saved for posterior viewing and interpretation. The system is then ready to perform the same operations all over again.

The following Flow Chart realizes the previous statements:

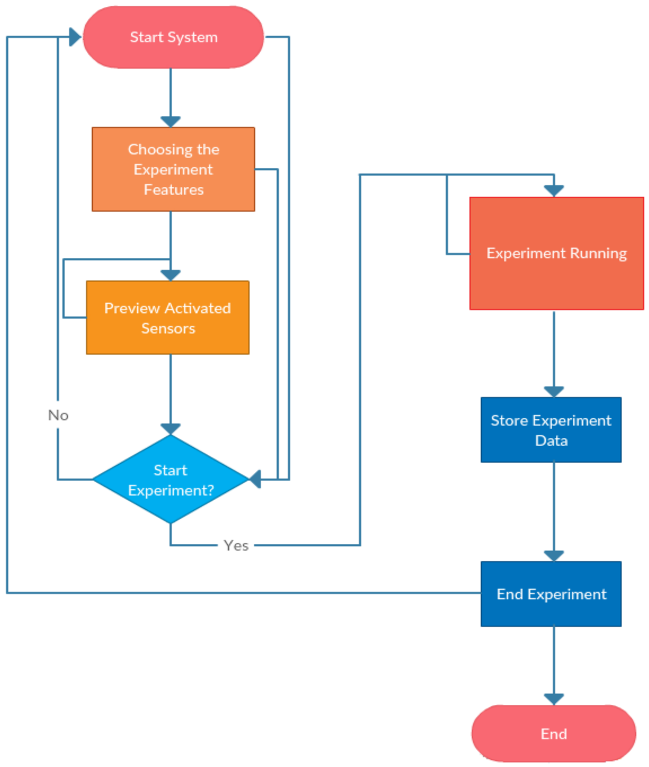

Figure 9 - General System Flow describing how to run an experiment

5.2 Data Synchronisation and Correctness

Two important aspects in terms of data synchronisation need to be verified. They are: 1). Correct Event Synchronization and 2). Correct acquisition and storage of ECG signal.

On the next sections we will deal with verifying those two points.

5.2.1 Event Synchronization

Making sure that all the signals are synchronised is important, since one of the requirements is that all the data must be correctly synchronised in order to allow us to relate and compare information between signals, allowing us to look in the right location and comparing the correct segment, and thus lowering the difficulty of presenting the various signals for inspection. We chose to use our experiment review application to show correct synchronisation since it crosses information between the events and signals, and if they are misaligned or with faults it would show visually. We performed a controlled experiment, where we used the Bitalino while recording the experiment. From time to time, we would introduce noise in the Bitalino, mark that noise with an event, while being recorded introducing that same noise. The result can be seen in Figure 10.

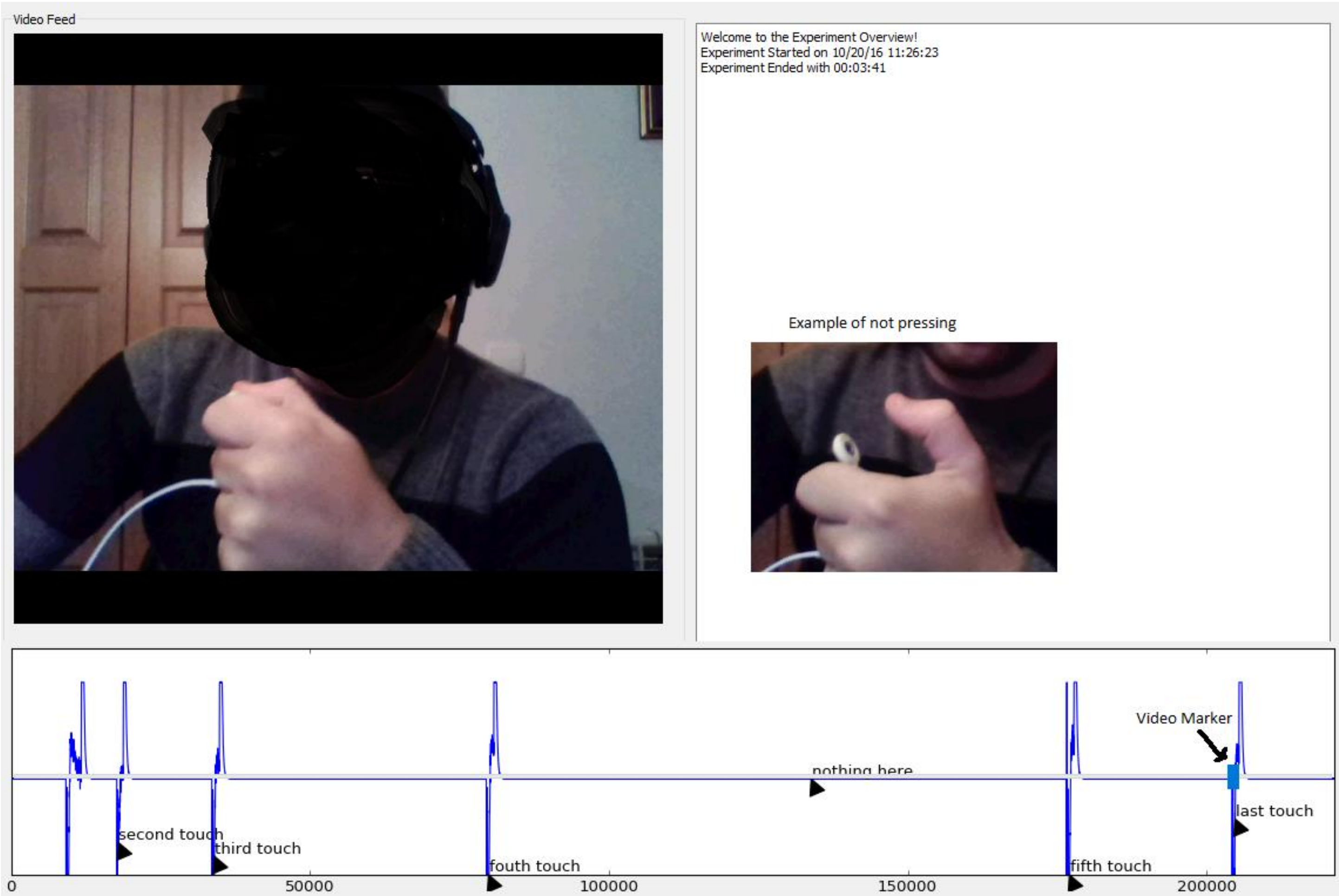

Figure 10 – PsyVidLab signal synchronization demonstration using a visual approach

Average delay between artifacts and corresponding event is 40ms. Delay between artifact and corresponding video action that created that artifact is 100ms.

Table 5 – Measurement delay between signals and events

Synchronization verification	Experiment duration = 30 seconds	Experiment duration = 5 minutes	Experiment duration = 30 minutes
Delay between artefact and event	Average: 40ms Standard Deviation: 3ms	Average: 50ms Standard Deviation: 6ms	Average: 48ms Standard Deviation: 5ms
Delay between artefact and video action	Average: 100ms Standard Deviation: 12ms	Average: 102ms Standard Deviation: 19ms	Average: 107ms Standard Deviation: 23ms

Table 5 shows delays between artefacts, events and video actions, quantifying it in terms of milliseconds. We considered the results shows acceptable.

In Figure 10 we can see that all the interference spikes have been successfully marked with events (the first spike was left unmarked on purpose) and on their correct locations. The video marker, which is used to tell and show us the correspondence in the video to the signal, clearly shows the noise being introduced by pressing the electrodes, and the resulting noise spike in the signal, all at the same time. All elements of the program are correctly synchronised, the video subsystem responsible in distributing frames, processing and recording (this includes the heart rate from video module) is synchronised with the ECG signal, and the event system, which is used not only to mark event on signals but also internally to control all the modules and subsystem in the program is also correctly synchronised. A snip showing how not pressing the electrode is like is also presented. We were able to measure the delay of the events in the program, that is, how long it takes to fire an event and that event being effectively active, which is less than 40ms on our system. We expect this number to change depending on the hardware.

5.2.2 Acquisition and Storage of ECG signal

In order to verify if the signal acquisition of the Bitalino module is functioning correctly, that is if no data is thrown away or lost, and that the recording and writing of the signals is in order and fully accounted for, we connected a ECG simulator to the Bilalino, and started and experiment where we only recorded the ECG data.

The following Figure illustrates the result:

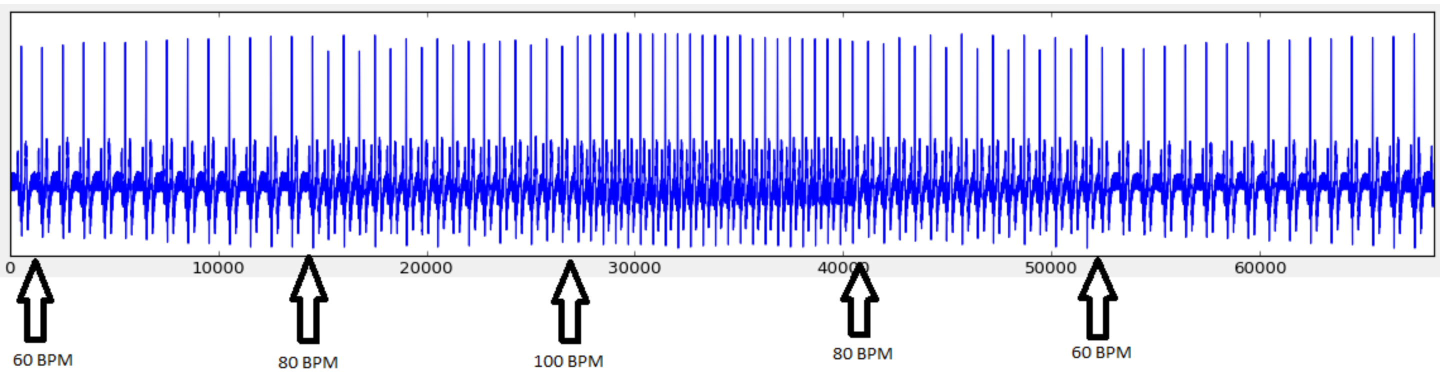

Figure 11 - PsyVidLab ECG validation

At the beginning we set the simulator at 60 BPM, then to 80 BPM, 100 BPM, 80 BPM and finally 60 BPM as seen from the previous Figure. No data or beats were missed, the waveform is complete and well structured, and the signal was saved and reloaded properly without any errors.

5.3 Implementation options

During the development of our system, several technical problems arose that ended up defining the nature of the system itself.

Initially when planning the development of the system, several major programing languages were considered for development, such as C++, Java, C# and Python. Since this system was to be used with a rich user interface, Java was put aside due to the limited capabilities of Swing (we are aware that alternatives exist such as Qt Jambi but was discarded due to having no experience at all on that framework). C++ and Python both featured interesting user interfaces, such as Qt for C++ and PyQt for Python, and it was later decided that Python was to be used since it was the language we were most comfortable with and with the most experience, including the PyQt framework. Python 2.7 and PyQt 4 were used to maintain compatibility with most libraries, such as Bitalino's and GoPro's. Qt was chosen because it features an inter-thread communication system based on events that allowed us to implement the system on a very simple and straightforward way. Plus, it disables the python's Global Interpreter Lock in favor of its own synchronization method.

Communication between threads is based on events, which are an unidirectional flow of information. These is a signal, an emitter, and a slot. The slot receives the signal that was emitted, and the emitter and the slot must be connected for them to be able to interact with each other. The signal can carry no information at all, and be used only as a flag or a warning, or it can be used to carry and move complex data structures between threads asynchronously. This method, which can be seen as an Asynchronous System Trap method, solved the inter-thread communication problems perfectly, preventing the occurrence of deadlocks

One of the features of our system is the capability of recording and viewing of biometric data in real time based on video. This means that several threads will attempt to access simultaneously the same hardware resource, in this case the camera. This is a typical multiple-consumer/one-producer problem. In order to solve it, a single thread was responsible to retrieve frames from the camera, store them in a queue by order, and send them to active sensors that required these frames.

Synchronization of the various signals is extremely important, so that we can compare the same time-frame between the various signals, recordings and stimulus. To solve this issue, first it was decided that all data from sensors should start recording at the same time, on the same instant. In order to accomplish this, sensors would report when they are ready through events, and when all the sensors are ready, the experiment and recordings would start. When it comes to terminating the experiment, all sensor threads would receive the command to terminate at the same time though events, and all data would be written to disk.

In order to display biometric signals in real time, we make use of line plots that change over-time. Python has a library called MatplotLib that takes care of all plotting needs, however it is not suited for real-time animation, since it takes a lot of time to build and render the graphs. This library is more suited for high quality graphics that are static. To solve this need, we made use of a different library called PyQtGraph, which has the capacity to render real-time graphs with thousands of points while supporting a toolbar for zooming in and out.

This system was to be used by stakeholders used to Windows. To address this situation, we developed our system in Windows, using the WinPython portable package which featured all the dependencies of our project, and could be moved between Windows Computers without installing any dependencies or any binaries.

When displaying video, we used the QT API called Phonon, which uses the windows media player backend as media source in Windows system, and the default media player in Linux systems. During our experiments, we came across a bug on this backend on windows that prevented large video files (greater than 1.5Gbs) from being correctly displayed from Phonon. The solution we found was to change the backend of Phonon manually to VLC, which had the requirement of having VLC installed on the system.

6 Pilot use and system validation

In this chapter we will discuss the early results we gathered from our experiments, in order to verify how our solution stands against both the state of the art and the reference experimental results provided by dedicated hardware, specifically Biopac and Bitalino.

We start with the controlled experiments results, in terms of average, standard deviation and average differences between estimated and reference heart rate time series, while displaying the normalised graph of both signals, and evaluating the trend of both time series. The second part of results evaluation is focused on a dataset in an uncontrolled environment, i.e., the setup was not designed with the purpose of this experiment, which will reinforce the analysis over a realistic dataset.

6.1 Analysis Method of experiments in controlled and uncontrolled scenarios

The following pipeline illustrates the method we used to calculate the results we have shown:

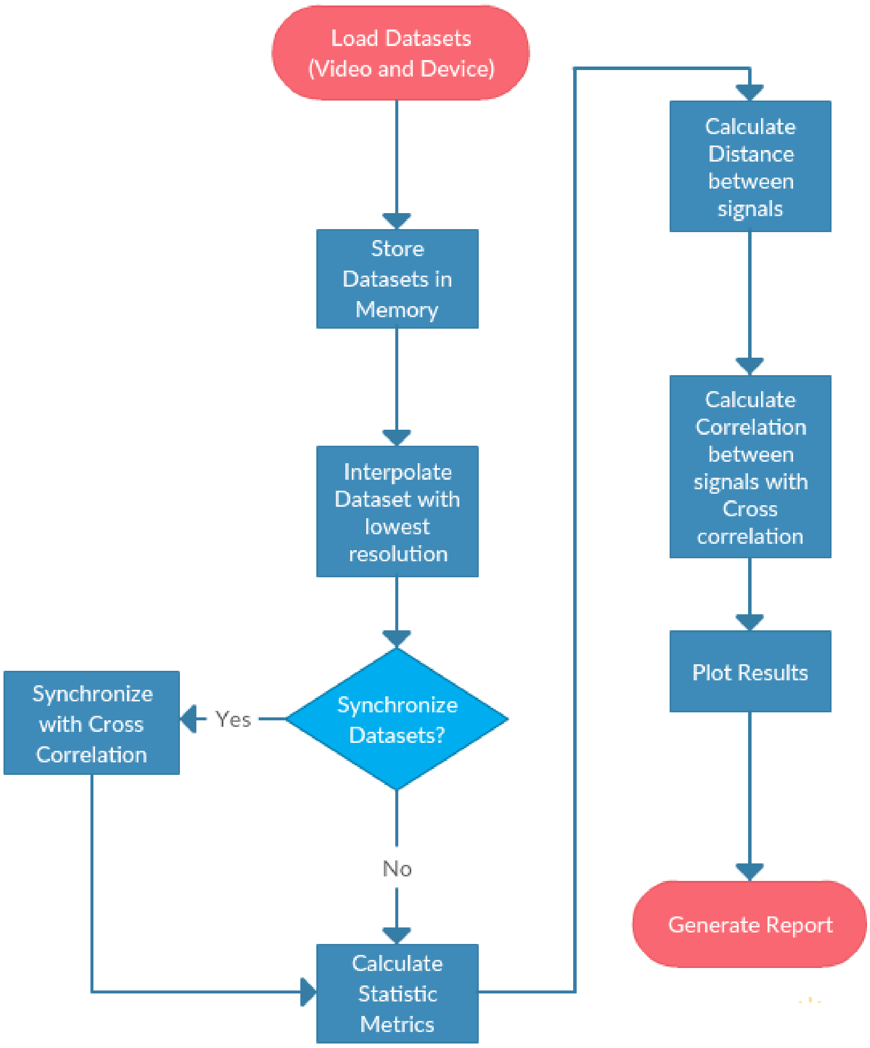

Figure 12 – Pipeline used to process the datasets from the experiments

The following bullets address the key stages of our pipeline:

- Interpolate Dataset with lowest resolution – Datasets don't necessarily have the same length. In order to accomplish that, we make use of interpolation, specifically interp1d from the scipy package for python, that interpolates a 1-dimensional function (we used linear interpolation). We tested the UnivariateSpline method in alternative, however we discarded it since this interpolation process created a wobbly pattern on segments that should be straight, altering the trend substantially.
- Synchronize Datasets with Cross Correlation – Sometimes, two datasets need to be synchronized if we don't know exactly if the start of one coincides with the start of the other. In this particular case, the ECG data provided by the psychologists carried no information at all of when did the readings started, all we know is that it started sometime before the stimulus video. We know exactly where the HR from video dataset started, which was right at the beginning of the stimulus video. In order to make sure that both datasets started at the same point in time, we decided to measure the Cross Correlation between the two using only part of both signals, that is, the baseline. This method would return the instance from which both signals would achieve the greatest correlation.
- Calculate correlation between signals – In order to perform this calculation, we used the Matlab function called xcorr to calculate the correlation coefficient between both time series with zero lag. A sliding window of 30 seconds was used to perform the calculation along the length of the time series.
- Calculate Statistic Metrics – In this part of our method, we calculated the average and standard deviation of the datasets.

Comparing the signal trend through correlation and amplitude of both signals is important, because it allows to verify if an artifact or event detected in one-time series, is present on the other time series, following the same trend, which may indicate that they share the same causal event. For example, a subject gets frightened by an image. This event should appear visible on the Heart Rate calculated from Biopac, usually in the form or a heart rate rise in amplitude. The heart rate signal estimated from video should show a similar rise, that is, the same tendency, in the same time frame.

6.2 Controlled Experiments to assess heart rate estimation

In this trial, we conducted three experiments featuring two heart rate signals, one estimated from video, and the other calculated from ECG with the Bitalino board. Three electrodes were used and placed on the subject's body as shown in Figure 13.

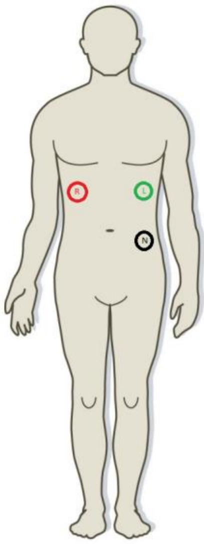

Figure 13 - Electrode positioning during experiments using Bitalino

The red mark refers to the right electrode, the green mark refers to the left electrode, and the black mark refers to the neutral electrode.

In this study, we designed the experiment and setup according to the literature. The user was asked to watch a film, which was composed by a baseline sequence featuring a peaceful tropical beach, followed by a sequence of funny clips. During the trials we marked the readings with events to verify if the various signals were synchronized, and recorded both ECG from Bitalino and HR from video to determine if these signals had correlation.

The experiment setup can be seen from the in Figure 13 and Figure 14.

Figure 14 - Experiment setup during application trials

We had two displays during our application trials, one was used to display the stimulus (in this case the laptop), and the other was used to view and monitor the biometric signals during the experiment in real time, as shown by figure 14. The Bitalino board and its electrodes are also visible in the figure.

Figure 15 -Stimulus was presentation during the application trials (in this case, a relaxing landscape).

Figure 15 shows how the stimulus video was presented to the subject, in full-screen, and with audio through head-phones.

The stimulus video featured a baseline of 3:14 seconds showing a relaxing tropical beach accompanied with relaxing music. The remaining 5:45 minutes of video featured an assortment of funny and short clips, making up 8:00 minutes of stimulus and biometric recordings.

Figure 16 represents the first application trial. As stated previously, we will show both signals together to facilitate visual interpretation of both signals in relation with each other.

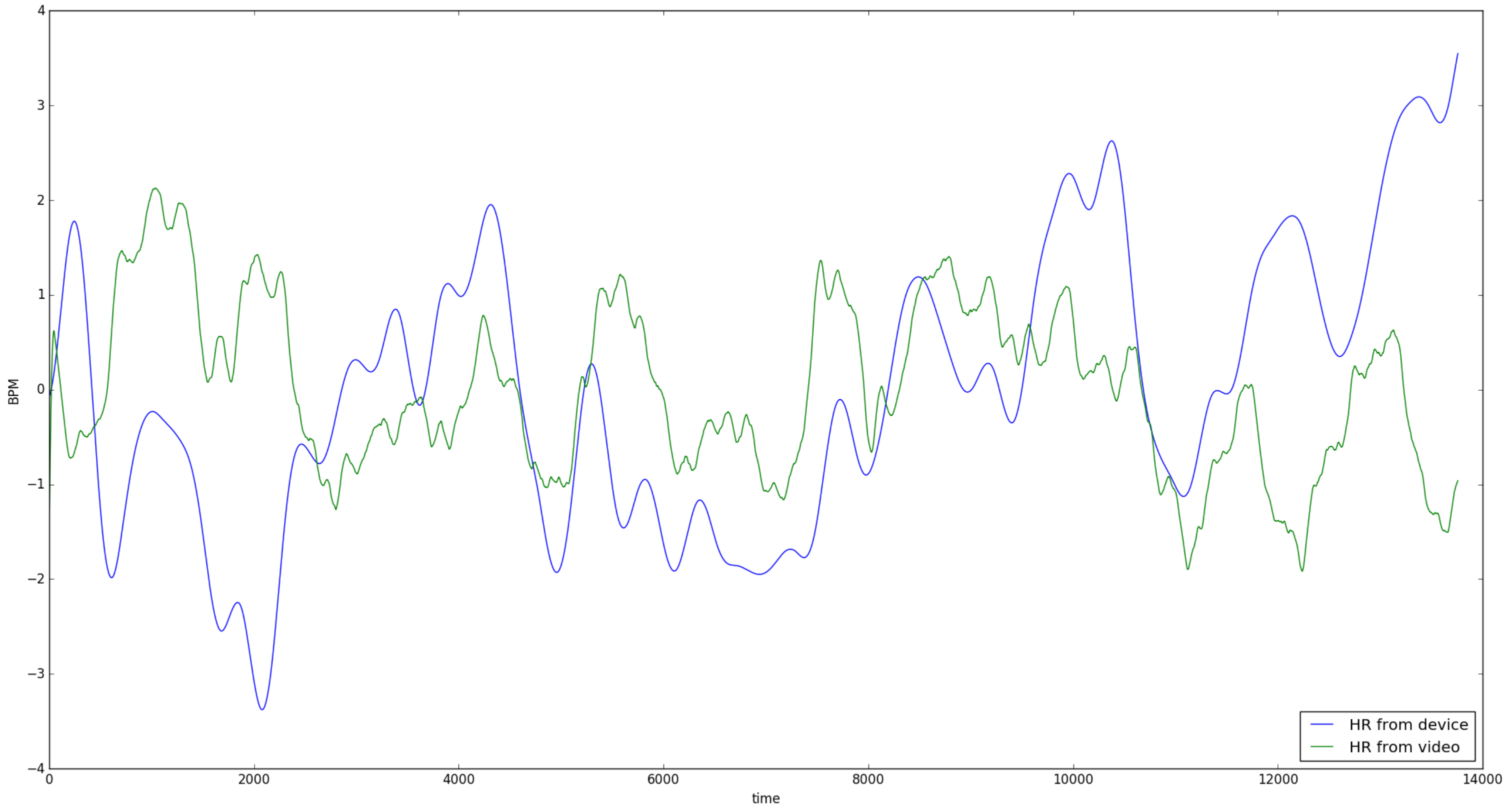

Figure 16 –Heart Rate from video and ECG comparison on Controlled Experiment 1 (without average)

Both signals appear to visually follow the same tendency, and the amplitude of the signals are quite similar on some regions. Figure 17 describes the correlation between the two-time series, in overlapping windows of 30 seconds, i.e., a running window of 30 seconds was used in order to calculate the correlation between the two-time series in short time intervals. In the contributions [33], [43] [44] the method used for HR estimation using video frames, is described as effective in short time intervals, therefore, in this dissertation we used the advisable time interval to validate the relation between the time series.

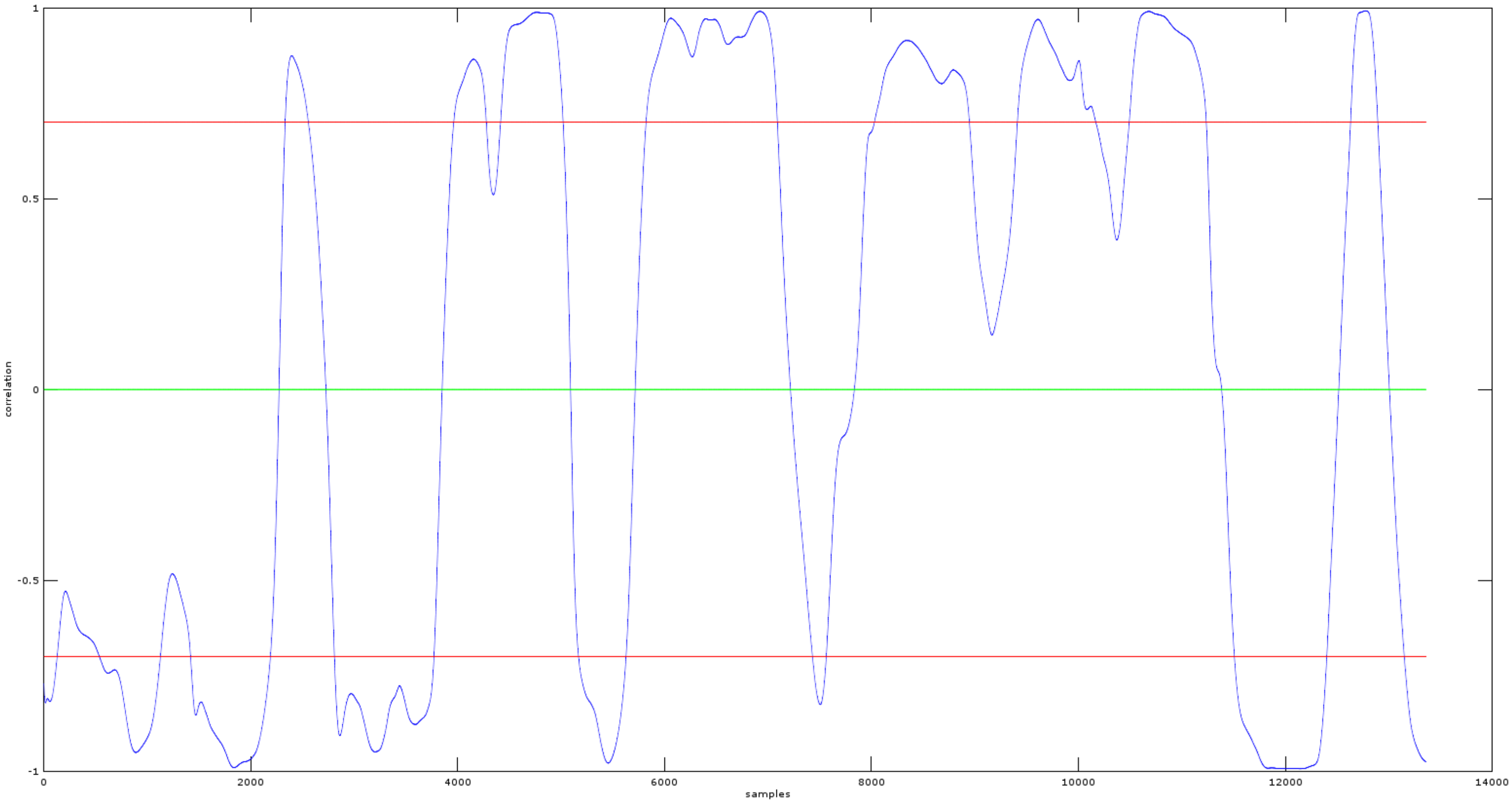

Figure 17 – Correlation over time between signals on Controlled Experiment 1

Figure 17 presents correlation intervals in which values are higher than 0.7, indicating a straightforward relation between the two-time series, which indicates that both have the same trend. There are also intervals in which the correlation values show negative correlation, in values lower than -0.7, which may indicate a possible delay between time series, because one is growing and other is decreasing. The correlation values inside the interval -0.7 and 0.7, are less in number than the in the other intervals, which indicate that the two-time series reveal almost in all intervals a relation.

Table 5 resumes the quantification of HR estimation comparison.

Table 6 – Controlled Experiment 1 results summary

Controlled Experiment 1	Heart Rate from ECG	Heart Rate from Video
Original Dataset Size	75	14309
Average (BPM)	97	92
Standard Deviation (BPM)	3	5
Average difference between signals	5	
Standard Deviation of average differences between signals	4	

The following table summarizes the results of the second controlled experiment.

Table 7 – Controlled experiment 2 results summary

Controlled Experiment 2	Heart Rate from ECG	Heart Rate from Video
Original Dataset Size	53	13928
Average (BPM)	69	92
Standard Deviation (BPM)	4	5
Average difference between signals	24	
Standard Deviation of average differences between signals	4	

Regarding the third controlled experiment, the results we gathered are shown in the following table.

Table 8 - Controlled experiment 3 results summary

Controlled Experiment 3	Heart Rate from ECG	Heart Rate from Video
Original Dataset Size	53	9471
Average (BPM)	66	86
Standard Deviation (BPM)	4	5
Average difference between signals	20	
Standard Deviation of average differences between signals	4	

After checking the previous tables, we verify that the average difference between signals is much greater on the second and third experiment than in the first. We are not certain what caused this increase in amplitude comparing with the first experiment, considering that the experiment circumstances were the same. However, the first participant is known to have high blood pressure and heart rate even at rest, usually oscillating around the 80 BPM at rest, while the other participants had no such ailments and oscillated around 60 BPM at rest. Further data should be collected in order to verify if the method is overestimating the HR in relation to the real values. Nevertheless, our main focus is to verify if the trend of the two HR estimations is aligned, because this may indicate that only an adjustment of the method is needed.

6.3 Experiments done on uncontrolled setup

In order to verify, and test our module, we accessed datasets from a few experiments conducted by psychologists that recorded the experiment in video and the electrocardiogram signal using Biopac. The past psychology experiments were labeled “Disgust”, because they aimed to stimulate disgust and measure the reactions of the subjects.

We were supplied with 28 psychology experiments with associated ECG readings. However, only two experiments were usable by us due to the fact that, in the remaining, the subjects face was either excessively rotated such that only half the face was visible, or the image circumstances were such that detection of the face was erratic and rare. To us, a good video had to feature at least 90% of face detection of video time. Otherwise the noise introduced could render the experiment useless.

In this study, data from two different participants were inspected. In the study, participants were asked to see a film. Each film presentation was preceded by a baseline period were a beach sunset with an acoustic guitar soundtrack was presented, after this interval the active film was presented during 25 minutes. The physiological channels were continuously sampled during the film presentations. The acquired ECG was sampled at 1000Hz, using the MP100 system and AcqKnowledge software (from Biopac Systems, Inc.). The electrodes were placed in the right hand, as well as in the right and left foot.

Figure 18 represents the Heart Rate from ECG and Heart Rate from video signals without average.

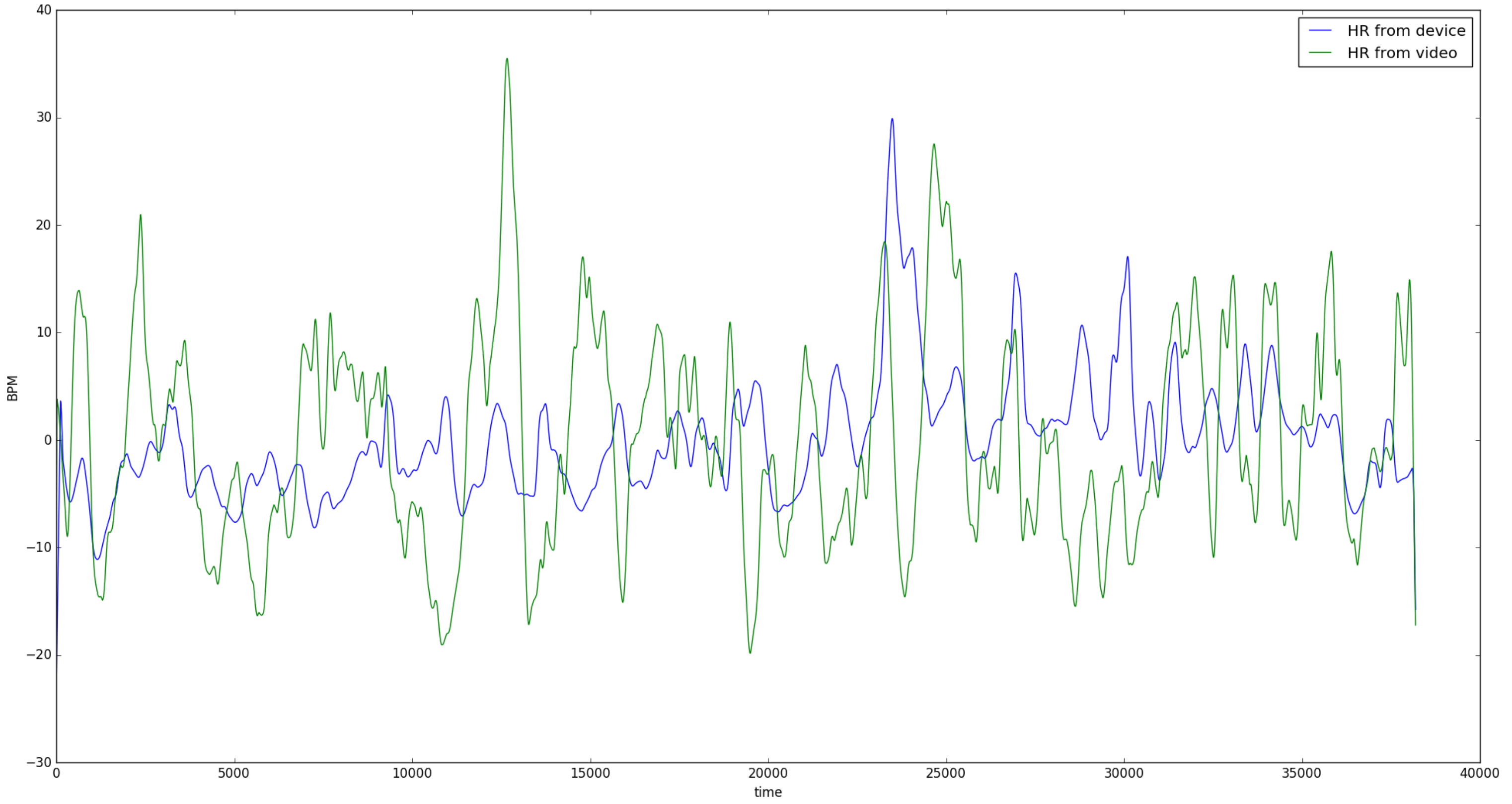

Figure 18 – Heart Rate from video and ECG comparison on first dataset (without average)

In Figure 18, the blue line refers to HR from Biopac, and the green line refers to HR from video. Since this data was collected in an uncontrolled environment the estimation is prone to more error. The visual inspection of the figure proves that, there is higher differences between the two HR series, even so there are intervals were the trend of the two-time series match, and also there is intervals were the time series behavior seems to be delayed (at approximately 2500 seconds).

The correlation in 30 seconds running windows was also calculated over this experiment. Figure 19 presents the obtained results.

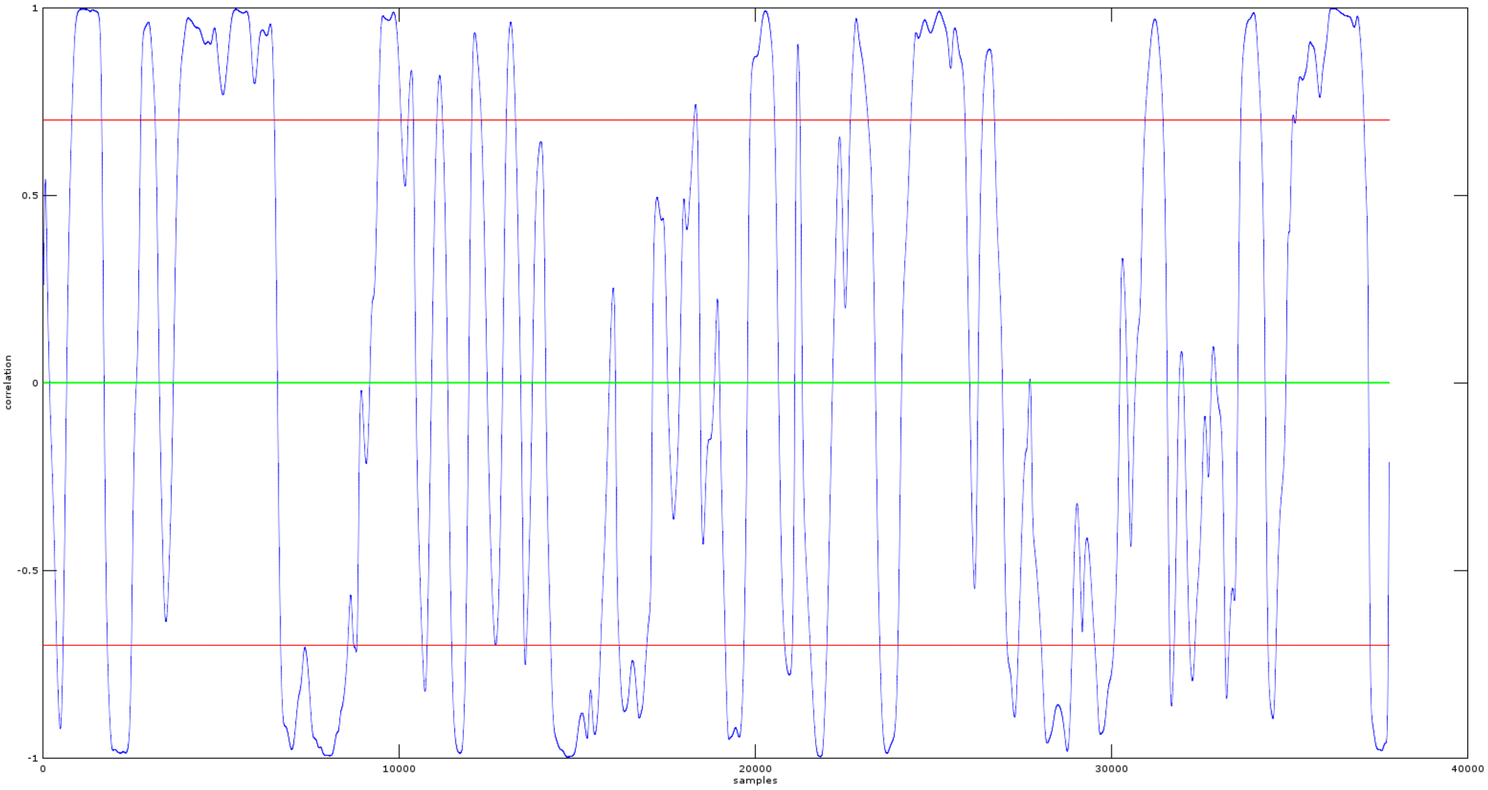

Figure 19 - Correlation over time between signals on first dataset given by the psychologists

Again, there are intervals were the match of the trend of the two-time series is similar (correlation values over 0.7). By the opposite, there is also intervals were the match is the opposite in terms of trend, which may indicate a possible delay (as also observed on Figure 17). In this example, there are several time intervals that lie within the >0.7 and the <-0.7 range, which means that is valid in the time frames (30 seconds) and can be used with some degree of confidence.

Table 8 summarises the results.

Table 9 – First dataset results summary

Past Dataset 1	Heart Rate from ECG	Heart Rate from Video
Original Dataset Size	3339	38324
Average BPM	70	73
Standard Deviation BPM	6	8
Average difference between signals	8	
Standard Deviation of average differences between signals	6	

The following table shows us the results obtained from processing both signals on the second dataset provided to us, and should give us a better insight about their relation, and help us in further analysis.

Table 10 – Second dataset results summary

Past Dataset 2	Heart Rate from ECG	Heart Rate from Video
Original Dataset Size	3967	38400
Average BPM	92	84
Standard Deviation BPM	5	7
Average difference between signals	9	
Standard Deviation of average differences between signals	6	

The average difference between the two-time series and their standard deviation are similar, despite the uncontrolled setting were the data was collected.

6.4 Results Summary

Table 11 - Comparison with state of the art methods with the same time frame (30 seconds)

Results Summary (in BPM)	Controlled Experiment 1	Controlled Experiment 2	Controlled Experiment 3	Syed Muhammad Imaduddin et al. [26]	Poh et al. [36]	Balakrishnan et al. [37]
Average Difference Between Signals	4	20	24	5	8	7
Standard Deviation	4	4	4	7	8	14

The algorithms reported on the contributions [33], [43] and [44] does not mention the number of experiments, also algorithms are offline, and operated on video of lengths of 30 seconds, while on our experiments, the videos length was 25 minutes in uncontrolled conditions, and 8 minutes on the Controlled Experiment Trial, analysis in real-time. The comparison between the literature and the controlled study may not be performed by a direct comparison. However, by the results inspection, we may infer that our results in short intervals are similar to the reported by these contributions (< 30 seconds segments).

In contributions [33], [43] [44] the used method for HR estimation from video is described to be used on highly controlled settings and also on short time intervals. In this study we decided to test the method on real settings, which may compromise the results. As may be observed on the results presented, we may infer that the method reveals promising results, near perfect correlation and comparable average signal difference, however, it will need to be adjusted to suppress all the noise and interferences presented on real settings, which may be performed by noise filtering or motion compensation.

These results help consolidate the validity and relevance not only of our work, but of this method as a means of acquiring heart rate without using intrusive approaches.

6.5 Emotion Detection Results

An emotion detection module was implemented; however, it was not integrated in the system. We based the method implementation on the state of the art [28]. The method produced interesting results. Figure 20 presents two emotion recognition results.

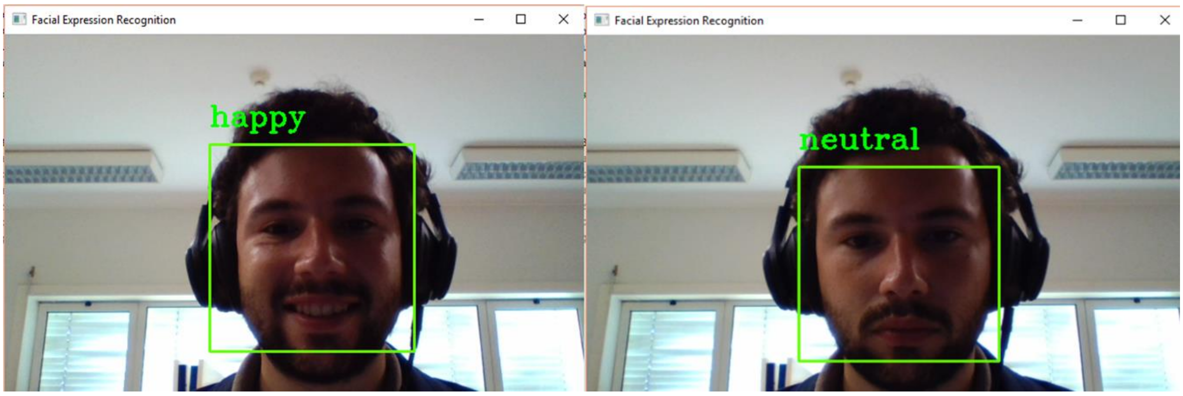

Figure 20 - Emotion Recognition module recognizing emotions in real time

The emotion recognition module uses a multi-layer neural network to recognize the emotion, and as such, this MLN requires training before running the system. To produce the results shown in the previous figure, we created a dataset with 5 images of a happy face and five images of a neutral face, and all the happy faces and neutral faces were detected successfully through the experiment (> 90% detection). However, when the subject showed other emotions, or when used on other subjects, the detector would fail to detect it and default to neutral. These results were in accordance with literature. The system was tested for different emotions, and the same results were achieved.

The available results are preliminary and require further training data. The ability to correctly recognize the emotion of a person depends heavily on the quality of the dataset; however, once trained, it is very fast, thus suitable for real time processing, making it a good choice to integrate with our system on future iterations.

7 Conclusions

We proposed PsyVidLab, an acquisition system to address requirements from psychology research experiments. PsyVidLab is modular and allows selection and acquisition from multiple sensors, such as heart rate from video, ECG from Bitalino, and video. It also has the ability to process recorded videos in order to extract heart rate from them using a standalone module.

PsyVidLab also has an application capable of displaying experiment results in a simple way, facilitating the analysis (it allowed us to verify the correctness and debug of certain components and modules of our system).

PsyVidLab was tested in 2 scenarios, specifically in video analysis and in real time experiment monitoring, with durations that greatly surpass previous trials (estimating HR for longer periods, in experiments with greater duration) made by contributions in the area. The heart rate from video allowed us to estimate heart rate from both recorded video and live video, and the Bitalino module provided us with a comparison for the live video experiments besides the inherent value it has by itself.

When it comes to the usefulness of the heart rate estimated from video, we need to ask two questions:

- Are the readings accurate enough (the results provided by the HR estimated from video module)?
- Can we accurately infer the correct physiological response from the signal?

Regarding the first question, since we had accurate results (average difference between signals inferior than 10 BPM) in our first scenario (previous dataset of recorded video) and in 1 out of 3 experiments on our controlled scenario, we can say that even though the algorithm that performs the estimation is not yet ready to be used due to lack of reliability (does not take in to account movement compensation and does not filter spurious heart beats), the potential is clearly there and the results are quite encouraging, especially considering the circumstances in which the readings were performed (in the case of the recorded videos scenario). The second question, which is related to the correlation between both signals, considering the early stage of development and simple algorithm, and the fact that when using time frames of 30 seconds sliding along both signals which produces good results (>0.7 and <-0.7 correlation coefficient) on many parts of the signals, we can say that the results are promising and we face the future of this technology with optimism. It is important to understand if it's possible though this method what factors impacted it the most and the possibility of past correlations becoming associated with events or experimental conditions that may bring added value to the quantification of the experiments.

The low values of correlation (< 0.7 coefficient) on some parts of the signals are most likely explained by the noise or interference introduced from motion. This motion does not necessarily need to come from the subject, which can be completely still, it may come from the face detection method itself. We noticed that the Viola-jones detector tends to introduce a lot of jitter even on subjects that are perfectly still, that is, the face coordinates oscillate, thus making the forehead region of interest shift and oscillate as well.

Recalling our initial objectives:

- Evaluate if the estimated heartrate has a good enough quality (relative to the reference heartrate, the estimated heart rate should follow the same trend and have, on average, a difference less than 10 beats per minute)
- Validate if the application acquires data correctly (specifically ECG), and the various signals are synchronized

For the first objective the average differences between time series fulfill this requirement as we could see from past tables that summarized the results, and are on par with other contributions results.

The second point has been discussed, and proven in section 5.2 Data Synchronization and Correctness to follow this requirement fully. The signals are shows to start at the same time, the events created by both the user and the application for internal usage are correct (with the right timings), and the video modules are synchronized with the signals since alterations on the signals are detected right on time on video. This behavior is shown by Figure 10.

Considering all these factors, our system manages to produce good results, with real-time performance allowing psychologists to perform their experiments in an ambulatory setup, customize it while providing all the features integrated in one place, with the option to present stimulus from the system itself, monitor all the sensors as the experiment goes on, and mark relevant events on the signals themselves. With further improvements in computer vision, specifically heart rate estimation from video, this technique is certain to improve and increase the potential of the system even further.

The correlation results between signals prove that our system and method of estimating heart rate from video provides the same good results such as the ones from the other contributions, in the same timeframe. Even though past contributions [33], [43] [44] don't show correlation along both signals, they claim that their results were optimal, on the same timeframe as them (30 seconds) we achieve good correlation values (> 0.7) therefor sporting good results as well. The method used does not take in to account motion compensation or noise filtering other than a rolling average. The focus was not in developing a method capable of performing better than all the contributions, but to include a method we believed was reasonable in a system to solve the needs of our stakeholders, and evaluate the correctness of our system against past contributions. It has been established that the results are promising, and thus this work will serve as a stepping stone, a proof of concept, for others to take and improve upon, so that eventually this system and method can be used in a real psychology scenario. For now, the seed of interest is planted, and its potential application in this field and others where contactless physiological monitoring is required, will definitely serve as motive to pursue further development and refinement.

7.1 Future work

Regarding the techniques and algorithms used, much is still needed to be done and improved. Instead of using python, future version of both applications should be written in C++ allowing increased performance and better manipulation of data-structures, plus solving the bottleneck of GIL (Global Interpreter Lock from python). Another aspect that should be improved would be the heart rate estimation from video algorithm itself. Borrowing inspiration from the method proposed by Balakrishnan et al. [44], the first detection of the face would be made with the Viola-Jones detector, and then track the face in subsequent frames with the Lucas-Kanade point tracker. This tweak should reduce the noise and interference caused by the Viola-Jones detector, possibly by quite a lot. Another upgrade that could help improve the quality of the signal would be the introduction of a NC-VT [45] filter to reject spurious heart beats at the end of the pipeline of the algorithm.

The user interface needs to be reworked and streamlined to facilitate human-computer interaction, as in its present form it resembles more of a technical preview than an actual system geared towards the final user. Several features that were planned could not be fully integrated, such as the emotion recognition module that works as a stand-alone application, and others that were not started due to them not being significant such as the gaze tracker module, should be fully integrated and tested on to the application.

But these techniques don't necessarily need to be "limited" to usage in psychology. These very methods could be used on any scenario where there is the need to extract heart rate information from previously recorded video, or monitor the heart rate on live scenarios where electrodes and probes are out of question, for example monitoring the heart rate of patients in a hospital emergency waiting room, airports, banks, and even on our homes. We believe this method of estimating heart rate from video will be capable of much better results in the future, further increasing the usefulness of these techniques. The promise and potential is certainly here; the work has just begun.

8 References

[1] – Ryan M. Niemiec & Stephan E. Schulenberg, “Understanding Death Attitudes: The Integration of Movies, Positive Psychology, and Meaning Management”, 2011, DOI: http://dx.doi.org/10.1080/07481187.2010.544517

[2] - Onur Asan, and Enid Montague, “Using video-based observation research methods in primary care health encounters to evaluate complex interactions”, 2015, DOI: 10.14236/jhi.v21i4.72

[3] - Gorodnichy, Dimitry, “Facial Recognition in Video”, 2003, http://nparc.cisti-icist.nrc-cnrc.gc.ca/eng/view/object/?id=684411b0-dbe8-49ea-9737-8699c593d13f

[4] - Daniela Palomba, Alessandro Angrilli, Alessio Mini, “Visual evoked potentials, heart rate responses and memory to emotional pictorial stimuli”, 1997, DOI: http://dx.doi.org/10.1016/S0167-8760(97)00751-4

{5] – Balcetis, Emily; Dunning, David, “See what you want to see: Motivational influences on visual perception”, 2006, DOI: http://dx.doi.org/10.1037/0022-3514.91.4.612

[6] - John E. Helzer et al, “Reliability of Psychiatric Diagnosis II. The Test/Retest Reliability of Diagnostic Classification”, 1977, DOI: 10.1001/archpsyc.1977.01770140026002.

[7] - Michael Strober et al, “Reliability of Psychiatric Diagnosis in Hospitalized Adolescents”, 1981, DOI: 10.1001/archpsyc.1981.01780270027002.

[8] – Open Science Collaboration, “Estimating the reproducibility of psychological science”, 2015, DOI: 10.1126/science.aac4716

[9] - J. Allen, “Photoplethysmography and its application in clinical physiological measurement,” Physiological measurement, vol. 28, no. 3, p. R1, 2007

[10] - F. Wieringa, F. Mastik, and A. Van der Steen, “Contactless multiple wavelength photoplethysmographic imaging: a first step toward camera technology,” Annals of biomedical engineering, vol. 33, no. 8, pp. 1034–1041, 2005

[11] - W. Verkruysse, L. O. Svaasand, and J. S. Nelson, “Remote plethysmographic imaging using ambient light,” Optics express, vol. 16, no. 26, pp. 21 434–21 445, 2008

[12] - M.-Z. Poh, D. J. McDuff, and R. W. Picard, “Non-contact, automated cardiac pulse measurements using video imaging and blind source separation,” Optics Express, vol. 18, no. 10, pp. 10 762–10 774, 2010

[13] - G. Cennini, J. Arguel, K. Aks¸it, and A. van Leest, “Heart rate monitoring via remote photoplethysmography with motion artifacts reduction,” Optics express, vol. 18, no. 5, pp. 4867–4875, 2010

[14] - Y. Sun, C. Papin, V. Azorin-Peris, R. Kalawsky, S. Greenwald, and S. Hu, “Use of ambient light in remote photoplethysmographic systems: comparison between a high-performance camera and a low-cost webcam,” Journal of biomedical optics, vol. 17, no. 3, pp. 0370 051–03 700510, 201

[15] - J. R. Estepp, E. B. Blackford, and C. M. Meier, “Recovering pulse rate during motion artifact with a multi-imager array for non-contact imaging photoplethysmography,” in Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on. IEEE, 2014, pp. 1462–1469

[16] - G. de Haan and A. van Leest, “Improved motion robustness of remote-PPG by using the blood volume pulse signature,” Physiological measurement, vol. 35, no. 9, p. 1913, 2014

[17] - H. E. Tasli, A. Gudi, and M. den Uyl, “Remote PPG based vital sign measurement using adaptive facial regions,” in International Conference on Image Processing, 2014

[18] - H. Zhu, Y. Zhao, and L. Dong, “Non-contact detection of cardiac rate based on visible light imaging device,” in SPIE Optical Engineering Applications. International Society for Optics and Photonics, 2012, pp. 849 806–849806

[19] - A. Chung, X. Y. Wang, R. Amelard, C. Scharfenberger, J. Leong, J. Kulinski, A. Wong, and D. A. Clausi, “High-resolution motioncompensated imaging photoplethysmography for remote heart rate monitoring,” in SPIE BiOS. International Society for Optics and Photonics, 2015, pp. 93 160A–93160A

[20] - R. Amelard, C. Scharfenberger, A. Wong, and D. A. Clausi, “Illumination-compensated non-contact imaging photoplethysmography via dual-mode temporally coded illumination,” in SPIE BiOS. International Society for Optics and Photonics, 2015, pp. 931 607–931 607

[21] – L. C. e Silva, T. Miyasato, R. Nakatsu, Dept. of Electr. Eng., Nat. Univ. of Singapore, Singapore “Facial emotion recognition using multi-modal information”, DOI: 10.1109/ICICS.1997.647126

[22] – N. Sebe, M. S. Lew, I. Cohen, A. Gard, T. S. Huang, Leiden Inst. of Adv. Computer. Sci., Netherlands, “Emotion recognition using a Cauchy Naive Bayes classifier”, DOI: 10.1109/ICPR.2002.1044578

[23] - Spiros V. Ioannou, Amaryllis T. Raouzaiou, Vasilis A. Tzouvaras, Theofilos P. Mailis, Kostas C. Karpouzis, Stefanos D. Kollias, School of Electrical and Computer Engineering, National Technical University of Athens, “Emotion recognition through facial expression analysis based on a neurofuzzy network”, http://dx.doi.org/10.1016/j.neunet.2005.03.004

[24] – Biopac, https://www.biopac.com/

[25] – Thearn, https://github.com/thearn/webcam-pulse-detector

[26] – Face Tracking and emotion detection results, http://vis-www.cs.umass.edu/lfw/results.html

[27] - Daniel J. McDuff, Member, IEEE, Justin R. Estepp, Member, IEEE, Alyssa M. Piasecki and Ethan B. Blackford4, Member, IEEE, MIT Media Lab, Cambridge, “A Survey of Remote Optical Photoplethysmographic Imaging Methods”, DOI: 10.1109/EMBC.2015.7319857

[28] - Michael Beyeler, Packt Publishing, “OpenCV with Python Blueprints”, chapter 7, ISBN 139781785282690

[29] – Bitalino, http://www.bitalino.com/

[30] – van Gent, P. (2016). Analyzing a Discrete Heart Rate Signal Using Python. *A tech blog about fun things with Python and embedded electronics.* Retrieved from: http://www.paulvangent.com/2016/03/15/analyzing-a-discrete-heart-rate-signal-using-python-part-1/

[31] – van Gent, P. (2016). Analyzing a Discrete Heart Rate Signal Using Python. *A tech blog about fun things with Python and embedded electronics.* Retrieved from: http://www.paulvangent.com/2016/03/21/analyzing-a-discrete-heart-rate-signal-using-python-part-2/

[32] – van Gent, P. (2016). Analyzing a Discrete Heart Rate Signal Using Python. *A tech blog about fun things with Python and embedded electronics.* Retrieved from: http://www.paulvangent.com/2016/03/30/analyzing-a-discrete-heart-rate-signal-using-python-part-3

[33] - Syed Muhammad Imaduddin, Yaseen Athar, Amir Ali Khan, Muhammad Murtaza Khan and Faisal Mahmood Kashif, School of Electrical Engineering and Computer Science, National University of Sciences and Technology (NUST), Islamabad, Pakistan and Google, California, USA, “A computationally efficient heart rate measurement system using video cameras”, DOI: 10.1109/ICET.2015.7389198

[34] - H.-Y. Wu, M. Rubinstein, E. Shih, J. V. Guttag, F. Durand, and W. T. Freeman, “Eulerian video magnification for revealing subtle changes in the world.” ACM Trans. Graph., vol. 31, no. 4, p. 65, 2012

[35] - W. G. Zijlstra, A. Buursma, and W. P. Meeuwsen-van der Roest, “Absorption spectra of human fetal and adult oxyhemoglobin, deoxyhemoglobin, carboxyhemoglobin, and methemoglobin,” Clinical Chemistry, vol. 37, no. 9, pp. 1633–1638, 1991

[36] - D. McDuff, S. Gontarek, and R. Picard, “Improvements in remote cardio-pulmonary measurement using a five band digital camera.” IEEE Transactions on Biomedical Engineering, vol. 61, no. 10, pp. 2593 – 2601, 2014

[37] - N. Garbey, Mac aand Sun, A. Merla, and I. Pavlidis, “Contact-free measurememnt of cardiac pulse based on the eanalysis of thermal imagery,” IEEE Transactions on Biomedical Engineering, vol. 54, no. 8, pp. 1418–1426, 2007

[38] - Jane Edwardsa, Henry J Jacksonb, Philippa E Pattisonc, “Emotion recognition via facial expression and affective prosody in schizophrenia: A methodological review”, 2002, DOI: http://dx.doi.org/10.1016/S0272-7358(02)00130-7

[39] - Singla, Rajneesh. "A new approach for mood detection via using principal component analysis and fisherface algorithm." Journal of Global Research in Computer Science 2.7 (2011): 88-92.

[40] - Thuseethan, S., and S. Kuhanesan. "Eigenface Based Recognition of Emotion Variant Faces." Available at SSRN 2752808 (2016).

[41] - van Gent, P. (2016). Emotion Recognition Using Facial Landmarks, Python, DLib and OpenCV. A tech blog about fun things with Python and embedded electronics. Retrieved from: http://www.paulvangent.com/2016/08/05/emotion-recognition-using-facial-landmarks/

[42] - Jönsson, P. (2007). "Respiratory sinus arrhythmia as a function of state anxiety in healthy individuals". International Journal of Psycho-physiology. 63 (1): 48–54. doi:10.1016/j.ijpsycho.2006.08.002

[43] - M.-Z. Poh, D. McDuff, and R. Picard, “Advancements in noncontact, multiparameter physiological measurements using a webcam,” Biomedical Engineering, IEEE Transactions on, vol. 58, no. 1, pp. 7–11, Jan 2011.

[44] - G. Balakrishnan, F. Durand, and J. Guttag, “Detecting pulse from head motions in video,” in Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, June 2013, pp. 3430–3437.

[45] - J. Vila, F. Palacios, J. Presedo, M. Fernandez-Delgado, P. Felix, and S. Barro, “Time-frequency analysis of heart-rate variability,” Engineering in Medicine and Biology Magazine, IEEE, vol. 16, no. 5, pp. 119–126, Sept 1997

[46] – emotient.com

[47] – affectiva.com

[48] – emovu.com